

nature



ANCIENT JEWELLERY

Pendants and human fossils represent earliest evidence of modern humans in Europe

Coronavirus

The race to resolve the protein structures of SARS-CoV-2

Electronic eyeball

An artificial visual system that mimics the human eye

Cellular landscape

The human body mapped to the level of single cells

Tackle coronavirus in vulnerable communities

The pandemic has hit care homes, prisons and low-income communities hardest. Researchers are ready to help, but need more data.

Respiratory pathogens spread like wildfire when people are in close contact. So it's little wonder that almost all of the 150 biggest coronavirus outbreaks in the United States have been in prisons, nursing homes, veterans' homes, psychiatric hospitals, meat-packing plants and homeless shelters, where people live or work side by side.

The phenomenon can be seen worldwide. Singapore seemed to have almost contained its epidemic until it became clear that the virus had been spreading undetected among migrant workers living in dormitories. And across Europe, homes for elderly people are among the worst hit.

Health officials are still failing to contain COVID-19 in shared spaces such as these because of the difficulties in achieving physical distancing. Measures, such as working from home, that protect healthier, wealthier and freer individuals, are often impossible to achieve for those whose jobs or accommodation make it impossible to self-isolate. Worse, there is little evidence to back up current policies intended to keep residents of communal spaces safe – or to support new ones.

Evidence-based strategies are urgently needed to prevent the spread of infection in shared settings, and to detect cases early. Researchers are ready to answer this call. But policymakers and health officials must first prioritize this research, and report data on caseloads and deaths so that epidemiologists can work out, in detail, what is going wrong. Most urgently, they must make regular testing available for high-risk groups, so that responders can intervene when cases first arise.

In many countries, testing is limited to people with symptoms such as a fever or severe cough, even though it is now established that infected individuals without symptoms can spread the disease. Asymptomatic cases can be particularly dangerous in communal spaces, where infections spread fast. In early April, for example, researchers testing people in a homeless shelter in Boston, Massachusetts, found that almost 90% of 147 people infected with the coronavirus did not have identifiable symptoms (T. P. Baggett *et al.* *J. Am. Med. Assoc.* <http://doi.org/ggts3>; 2020).

Analyses of outbreaks in US nursing homes and in prisons have found that more than half of infected residents and staff didn't show obvious symptoms at the time of testing. Some epidemiologists, geneticists and social scientists are rightly urging policymakers to change the testing criteria

“There is little evidence to back up current policies intended to keep residents of communal spaces safe.”

so that people in communal settings are tested regularly, regardless of whether they have symptoms.

With more data, epidemiologists would be able to evaluate and compare interventions to see which work best. For example, are face masks preventing transmission? How effective is the practice of positioning beds in homeless shelters two metres apart? Would it be safer for homeless people to be accommodated outdoors, in tents, if single-occupancy accommodation isn't available?

In addition, by sequencing the viruses spreading in a facility, researchers can determine how often people are introducing viruses from outside, and to what extent infections are being amplified in communities. And full cost analyses will help policymakers to compare the total costs of different solutions. Something that seems expensive upfront might, over time, result in lower overall costs once other expenses, such as hospital stays, are factored in.

In the United States, which still has the world's highest number of confirmed deaths from COVID-19, scientists are ready to do more. Several academic labs say they can run thousands more tests than they are currently processing, and some have developed easier-to-deploy tests. For example, on 8 May, the US Food and Drug Administration permitted the emergency use of a test based on the gene-editing tool CRISPR that can be processed using less-sophisticated equipment than is required for many other tests.

But, for researchers to be more involved, they must be integrated into state-wide testing strategies that link them to health departments. And these agencies must, in turn, be prepared to respond to positive diagnoses.

At the moment, that is not a given. Alarming, some researchers have told *Nature* that officials are reluctant to survey people in communal spaces, because infected individuals will then need to be isolated, and their contacts potentially tested and quarantined, too. This could, in turn, mean providing housing, or paying wages to quarantined essential workers. These are difficult and expensive interventions, but ignoring the problem will not make it go away.

Wanted: accurate reporting

A lack of transparency is another obstacle to epidemiological analyses. According to the US Centers for Disease Control and Prevention, 30% of jurisdictions aren't reporting COVID-19 cases in prisons as a separate, identifiable category. Some jails are reporting outbreaks as a single event, rather than listing the number of cases. And many state public-health departments aren't reporting infections and deaths among residents of homeless shelters and nursing homes. An outbreak at one nursing home in New Jersey was discovered only when police found 17 dead bodies piled up inside.

This cannot continue. Facilities should report what's happening within their walls, and states should make anonymized data available quickly.

Some cities provide a model for others to follow. In Seattle, Washington – where the first US COVID-19 outbreak was detected – the public-health department has an online dashboard devoted to reporting daily cases and deaths in care homes. The city's partnership between these

facilities, researchers and the public-health department helped to reduce new COVID-19 cases in care homes from 748 in March to 72 in the first 2 weeks of May.

The lack of action elsewhere is an outrage. It isn't getting the attention it deserves because the people who are most affected are those least able to make their voices heard. Those who are poor, from minority communities, elderly, incarcerated, chronically ill or homeless are among the most marginalized in society. Their needs have been ignored in part because they have less access to policymakers. But they should not need to make their case – those in power should already be paying attention.

Researchers, however, can do their part. They understand the need to curb this pandemic among the most vulnerable people, and must make sure they work with these groups to study the pandemic and to analyse and highlight its devastating impacts. Policymakers must act on what they find. Until countries beat this disease in the places hit hardest, they won't be able to beat it at all.

Everyone wins when patents are pooled

The spirit of collaboration is being tested as vaccine development gets under way.

Last week, the leaders of Ghana, Pakistan, Senegal and South Africa co-signed an open letter urging that research and intellectual property on coronavirus vaccines be shared freely – and that vaccines be distributed fairly – so that the poorest countries do not lose out. It is unfortunate that such a letter needed to be written in the middle of the worst pandemic in decades. But it was unavoidable, because some governments – including those funding the first wave of research and clinical trials – have not yet committed to the principles of fully open science and innovation.

This contrasts sharply with the rapid sharing of findings and expertise among researchers being reported daily. In a Feature on page 252, we cover one example of such collaboration. Since January, researchers have been working across the globe and around the clock to reveal the structures of key proteins that make up the new coronavirus. Their achievements are the result of free-flowing exchange between university laboratories and national synchrotron facilities in countries including China, Germany, the United Kingdom and the United States. Work that would normally have taken months – or even years – has been completed in weeks. But rather than building on this cooperation, some countries are retreating into a kind of techno-protectionism, which serves neither science nor society.

On 10 January, when researchers in China and Australia shared the genome sequence of SARS-CoV-2 (F. Wu *et al.* *Nature* 579, 265–269; 2020) online, a global network of

biologists interested in the structure of viral proteins set to work. The network included the Center for Structural Genomics of Infectious Diseases, a consortium of 40 scientists across 8 institutions in the United States and Canada, which played a central part in the project.

Top of the consortium's to-do list was to plan which proteins to tackle first, and which lab would take on which protein. The teams then set about getting high-resolution snapshots of these proteins, which enable the virus to enter cells and replicate. Thanks to this work and similar efforts elsewhere, there are now more than 170 structures of whole or partial proteins alone or bound to a drug or receptor. The visualizations generated by this work can be used to find ways to neutralize the virus with drugs or vaccines.

Simultaneously, structural biologists at ShanghaiTech University in China began the task of revealing the structure of a key enzyme, M^{pro}, that the virus needs to replicate. Work that needed two months for SARS-CoV, the virus that caused the outbreak of severe acute respiratory syndrome (SARS) in 2003, this time took just one week. The team deposited its results in the Protein Data Bank – an open-access digital repository for 3D biological structures – ready for researchers around the world to access. As they worked, Shanghai team members collaborated with structural biologists at the University of Oxford, UK, to share knowledge and avoid overlap.

But when it comes to distributing some of the fruits of that knowledge, this spirit of cooperation looks to be at risk. It is crucial that any vaccine, once proved to work, can be made and distributed quickly in every country. For this to happen, the holders of intellectual property must pool their know-how so that companies large and small can participate in this emergency effort. Intellectual-property sharing initiatives are under way, but, as *Nature* went to press, neither the US nor UK governments seemed ready to support these efforts. This is unacceptable during a pandemic, when lives are at stake and the world's population needs to be immunized. The research that has got us to this point has been pooled, and governments are funding the vaccine effort. For these reasons, intellectual property has to be shared.

Patent pooling is not simple, but there's a wealth of literature from life-sciences patent law and case studies from the field of development studies that can help to make it work. And there is an important principle at stake. There is little justice, as economist Mariana Mazzucato at University College London often argues, if citizens have to bear many of the financial risks in such an endeavour, but most of the profits go to a small group of companies (and possibly a few universities) once a vaccine is ready to be rolled out.

Scientists are not exempt from competition: the race to publish a paper or patent a molecule is all too common. But in the race to solve the structure of SARS-CoV-2, the competitors have mostly worked together and shared credit – and that is how they, and the hundreds of researchers working in complementary fields, must continue as vaccines and drugs move into clinical trials. It is a tribute to the scientists involved that they immediately understood that a pandemic requires a different way of working. It is a tragedy that some governments do not.

“**Scientists immediately understood that a pandemic requires a different way of working.**”

World view



By Harriet A. Washington

How environmental racism fuels pandemics

Toxic living conditions inflate COVID-19 death rates. Scientists must track how and why.

As data accumulate that, in some places, people of colour are much more likely than white people to become ill and die of COVID-19, more discussions are grasping at factors beyond race to explain why. I'm used to this. When my book on environmental racism came out last year, one of the most common questions I received concerned alternative explanations for the greater ill health of minority ethnic groups. Surely, I was asked, the issue is not race, but poverty?

Poverty is a risk factor for becoming unwell. But racial disparities in exposure to environmental pollutants are greater factors that remain even after controlling for income. African Americans who earn US\$50,000–60,000 annually – solidly middle class – are exposed to much higher levels of industrial chemicals, air pollution and poisonous heavy metals, as well as pathogens, than are profoundly poor white people with annual incomes of \$10,000. The disparity exists across both urban and rural areas.

We need to take a longer, harder look at environmental racism – systems that produce and perpetuate inequalities in exposure to environmental pollutants. These can persist even in the absence of malevolent actors. The main culprits include indifference and ignorance, inadequate testing of industrial chemicals, racism, housing discrimination, corporate greed and lax legislation from, in the United States, a weakened Environmental Protection Agency. To combat these, society must actively take responsibility. By anticipating the outsized environmental assaults that people of colour face, we can act to protect lives during the current pandemic and future outbreaks.

It's true that pathogens are democratic by nature. It's also true that marginalized minority ethnic groups have increased exposure to environmental pollution and reduced access to health care. All this creates physical and social vulnerabilities that leave people of colour less able to resist and survive infections such as the coronavirus. This is not only a problem in the United States. In April, the UK Intensive Care National Audit and Research Centre estimated that 35% of people in intensive care with COVID-19 are black, Asian or members of other minority ethnic groups, nearly triple their proportion in the UK population. The first ten physicians in the United Kingdom known to have died from COVID-19 were also from black, Asian or minority ethnic groups. Crowded housing and working conditions have been suggested as a reason for the disparity. Only 2% of white people in the United Kingdom live in crowded conditions, but 30% of Bangladeshi, 16% of Pakistani and 15% of black African households are overcrowded. Black and minority

We need to look beyond blame-shifting assumptions that genetic differences or lifestyle 'choices' can explain medical disparities."

Harriet A. Washington is the author of several books, most recently *A Terrible Thing To Waste: Environmental Racism And Its Assault On The American Mind*. e-mail: hw2140@columbia.edu

ethnic people are also more likely to live in 'deprived' areas that are closer to sources of industrial pollution – from lead-tainted water in Flint, Michigan, to nerve gas, arsenic and polychlorinated biphenyls in Anniston, Alabama.

Black and minority ethnic populations are also more likely to live in neighbourhoods where they are exposed to high levels of lead and to air pollution. Greater exposure to air pollution has long been tied to shorter life expectancy. It can exacerbate heart diseases, trigger hypertension and compromise immune systems. A preliminary study published on a preprint server (Wu, X. *et al.* Preprint at medRxiv <https://doi.org/10/ggrpcj>; 2020) linked exposure to an increased likelihood of dying from COVID-19. Poor access to nutritious food makes matters worse. The term food deserts is often used for neighbourhoods that lack grocery stores and other vendors of fresh produce. I prefer 'food swamp' because such neighbourhoods are often teeming with places selling junk food, alcohol and tobacco. This foments obesity and nutritional deficiencies, which magnify the harms of environmental pollution. Vitamin C, calcium, and iron in the diet prevent the absorption of lead, a poisonous metal. A similar argument can be made about access to green spaces and exercise facilities.

To mitigate and prevent these and other inequalities, we need to collect and disseminate better data. Authorities need to document race, and not assume relevant information can be captured by socio-economic status. The US Centers for Disease Control and Prevention agreed to report deaths by race and ethnicity only when put under pressure, and several weeks after the first US death. As of 16 May, only 40 states were reporting the race of people who had died of COVID-19, and only 3 were reporting race for people who received COVID-19 tests.

In the 1990s, health scholars and practitioners were caught unaware by the higher rates of HIV infection and mortality in communities of colour. The pattern is being repeated in deaths from COVID-19. Policymakers need to recognize this and target interventions – perhaps by increasing testing for vulnerable populations. We need to look beyond blame-shifting assumptions that genetic differences or lifestyle 'choices' can explain medical disparities. Risks such as where a person lives or what they eat often reflect realities that lie beyond that person's control. Social distancing is impossible for someone who lives in a crowded flat and must work cheek by jowl in a meat-packing plant.

Instead, treatment and research must be designed using information about differences in access to health care and exposure. This approach will help everyone: as Tedros Adhanom Ghebreyesus, head of the World Health Organization, says: "No one is safe until everyone is safe." Most of all, we must remember that if we don't confront environmental racism directly, we cannot overcome it.

News in brief



DOGS CAUGHT CORONAVIRUS FROM THEIR OWNERS, GENETIC ANALYSIS SUGGESTS

The first two dogs reported to have coronavirus probably caught the infection from their owners, say researchers who studied the animals and members of the infected households in Hong Kong. An analysis of viral genetic sequences from the dogs showed them to be identical to those in the infected people.

Researchers suspected that the infection had been passed from the owners to the dogs, and the direct genomic link strongly supports that, says Malik Peiris, a virologist at the University of Hong Kong who led the study, which was published in *Nature* (T. H. C. Sit *et al.* *Nature* <http://doi.org/dvt4>; 2020).

The study showed no evidence that dogs can pass the infection to other dogs or to people, but it is impossible to be certain in which direction the virus travelled “so we have to keep an open mind”, says Peiris.

Although the analysis confirms that people with COVID-19 can infect dogs, the probability of this happening is low, says Arjan Stegeman, a veterinary epidemiologist

at Utrecht University in the Netherlands. In the study, only 2 of the 15 dogs who lived with infected people got the disease.

Since the infections in the two canines in Hong Kong — a Pomeranian and a German shepherd — were reported, other pets have tested positive for the SARS-CoV-2 virus, including a cat in Hong Kong and another two in New York state. Four tigers and three lions at New York City's Bronx Zoo also tested positive. Studies in cats have found that they can pass the virus to other felines.

The Hong Kong study detected viral RNA and antibodies in both dogs, and live virus in one of them. Neither dog became noticeably sick.

The findings support the results of an April study, in which researchers in China infected dogs with SARS-CoV-2, says Thomas Mettenleiter, a virologist at the Federal Research Institute for Animal Health in Riems, Germany. Dog owners who test positive for the coronavirus should be cautious when handling their pets, he says.

PUBLISHERS UNITE TO TACKLE ALTERED IMAGES

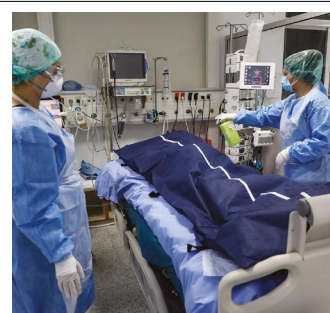
The world's largest science publishers are teaming up to establish standards for catching suspicious images in research papers. A new working group — the first formal cross-industry initiative to discuss the issue — aims to set standards for software that screens papers for altered or duplicated images during peer review.

Journal editors have long been concerned about how best to spot altered images, which can result from honest mistakes or efforts to improve the appearance of images, as well as from misconduct. So far, most journals haven't employed image-checkers to screen manuscripts, saying that it is too expensive or time-consuming; and software that can screen papers on a large scale hasn't been available.

The new cross-publisher working group aims to lay out minimal requirements for software that spots problems with images, and to look at how publishers could use the technology across hundreds of thousands — or even millions — of papers.

The group began meeting in April, having been set up by the standards and technology committee of the STM, a global trade association for publishers, based in Oxford, UK. It includes representatives from publishers including Elsevier, Wiley, Springer Nature and Taylor & Francis.

“The ultimate goal is to have an environment that helps us, in an automated way, to identify image alterations,” says the group's chair, IJsbrand Jan Aalbersberg, who is head of research integrity at Elsevier.



CORONAVIRUS HINDERS AUTOPSIES, DEPRIVING RESEARCH OF CRUCIAL TISSUE

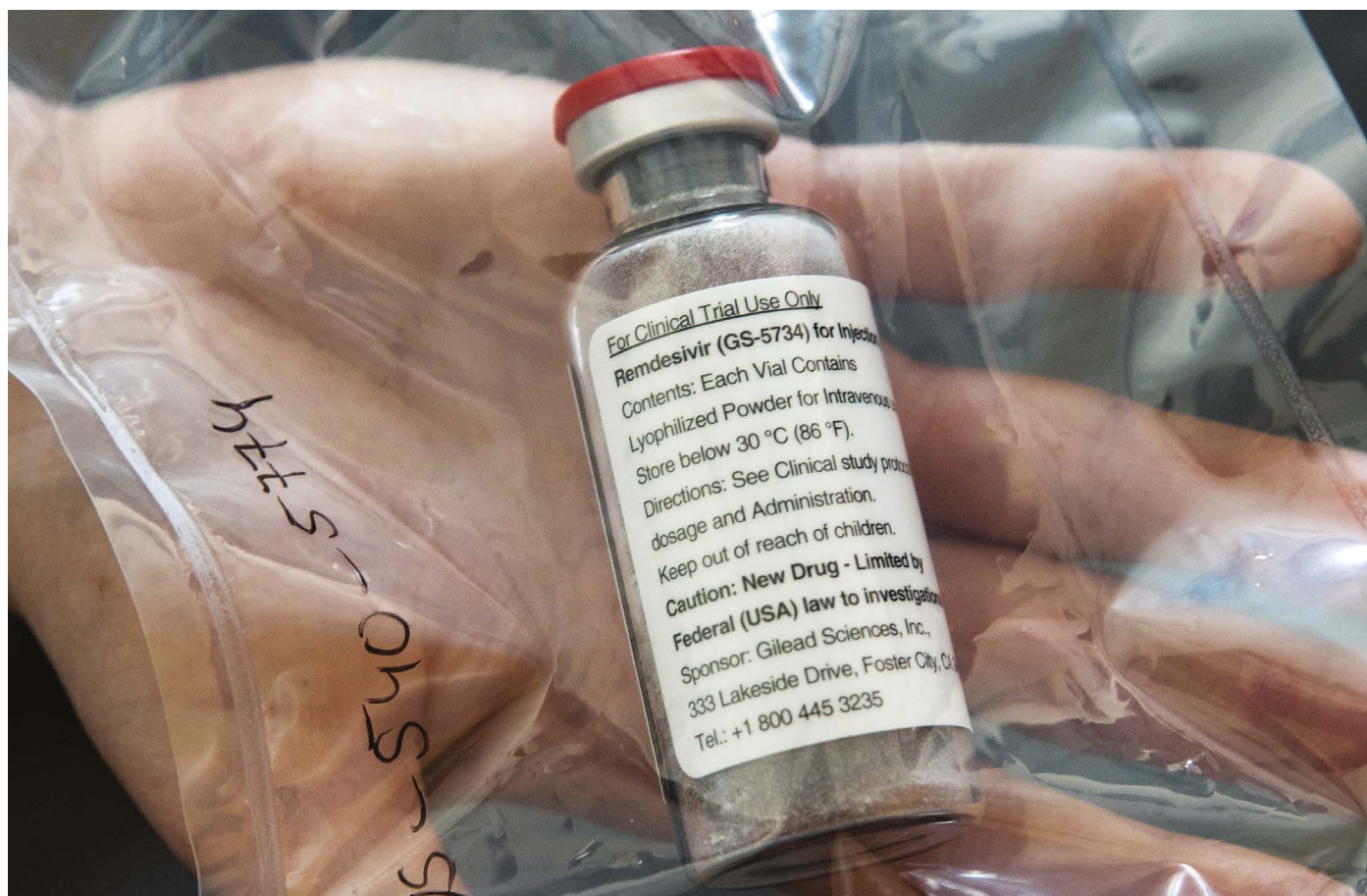
As researchers worldwide struggle to understand COVID-19's effects on the body, they are clamouring for tissue samples from patients. But the raging pandemic and ongoing lockdowns have complicated efforts to do autopsies and collect the tissue needed to understand how the coronavirus attacks organs including the lungs, heart and brain.

Autopsies are always painstaking work, but the pandemic means that health-care systems are overwhelmed, protective equipment is in short supply and pathologists are at high risk of infection.

But some researchers have found ways around the obstacles. Pathologist Marisa Dolnikoff at the University of São Paulo and her colleagues have been performing minimally invasive autopsies using needle biopsies to understand why some patients develop blood clots.

Researchers now want to collect and share such samples and results systematically. A team of pathologists including Roberto Salgado at the GZA-ZNA Hospitals in Antwerp, Belgium, is creating a global COVID-19 pathology repository. The group is working with the World Health Organization to create guidelines for the safe collection of autopsy samples and a standardized way of recording the results.

News in focus



ULRICH PERREY/POOL/AFP/GETTY

Production of remdesivir, an antiviral drug approved to treat COVID-19, is ramping up.

DOZENS OF CORONAVIRUS DRUGS ARE IN DEVELOPMENT — WHAT HAPPENS NEXT?

Drug-makers face supply-chain weaknesses and sourcing issues as they ramp up complex production processes to meet global demand.

By Heidi Ledford

The world was waiting for any sign of hope in countering the COVID-19 pandemic when researchers released the first encouraging data from a large clinical trial of the antiviral drug remdesivir last month. The drug, they said, reduced the time to recovery from COVID-19 by a few days – not enough to be branded a ‘cure’, but enough, it’s hoped, to relieve some pressure on overwhelmed health systems.

The discovery of remdesivir’s potential

focused attention on the next problem facing the development of COVID-19 therapeutics: ramping up complex manufacturing processes to address a global pandemic. It is likely to be one of the biggest drug-making challenges the world has ever faced. Some of the therapies being tested against COVID-19 are new and difficult to produce. Others – even if they are relatively simple compounds that have been in use for decades – face complications such as supply-chain weaknesses as drug-makers try to scale up production.

“A major rate-limiting step is going to be

manufacturing,” says Ezekiel Emanuel, a bioethicist at the University of Pennsylvania in Philadelphia. “Getting up to hundreds of millions of doses is hard.”

Spectrum of drugs

Researchers are working furiously to test a wide variety of potential COVID-19 treatments. Those therapies run the gamut of complexity, from familiar generic medications, such as the malaria drug hydroxychloroquine, to experimental small molecules such as remdesivir, which was previously trialled against

the Ebola virus. Scientists are also exploring antibody treatments that tamp down the body's immune response when it becomes destructive, which happens in some people critically ill with the coronavirus. And if the history of infectious disease is any guide, it will take a combination of drugs – each with a distinct, even if relatively minor, impact on the disease – to tame the new coronavirus.

Each treatment will face different challenges when scaling up production, says Stephen Chick, who studies health-care management at the business school INSEAD in Fontainebleau, France. “If it's successful and the technology is then adopted, you need to be prepared to deliver,” he says. “And if you're not, you're in trouble.”

Remdesivir's maker, Gilead Sciences in Foster City, California, has been working for months to scale up production of the compound, even before the latest data release. After the US Food and Drug Administration authorized use of the drug for COVID-19 under emergency rules on 1 May, the company announced that it had reached out to drug manufacturers around the world to find ways of boosting production.

By then, Gilead had already been streamlining its manufacturing process – reducing the time needed to produce large batches of the drug from 9–12 months to 6–8 months – and searching for alternative sources for the rare chemicals needed to make it. The company has projected that it could make enough remdesivir to treat one million people by the end of the year, and potentially twice as many if it finds that lower doses of the drug are sufficient to reduce recovery time from COVID-19.

But it also warned that production of remdesivir relies on a complex chemical synthesis – with individual steps that can take weeks to perform – and could be derailed by shortages of key ingredients. Remdesivir's structure is similar to the nucleotide building blocks the virus uses to copy its RNA genome. By imitating those building blocks, remdesivir blocks the enzyme that the coronavirus uses to replicate itself.

Bargain search

Gilead faces a particular challenge because it was not making large amounts of the drug when the pandemic started. But even for compounds that are already produced in bulk – such as hydroxychloroquine and its chemical cousin chloroquine – scaling up presents a problem, says David Simchi-Levi, an operations researcher at the Massachusetts Institute of Technology in Cambridge.

Over the past two decades, manufacturers in many industries have been shifting to a ‘lean’ manufacturing model that reduces the amount of raw materials and finished product they keep in stock. “This was successful in terms of reducing costs,” Simchi-Levi says.

“But it increased exposure to risk.”

In addition, companies have been seeking low-cost suppliers of raw materials in countries including China and India. When a crisis such as a pandemic strikes, those countries might clamp down on exports of pharmaceutical ingredients to ensure availability to their own people.

Simchi-Levi and his colleagues' research in the automotive industry showed that the riskiest links in the supply chain were providers of crucial components that cost as little as 10 cents. The same could be true of other industries, he says, including pharmaceuticals, where there are already concerns about having enough glass vials to produce and distribute a vaccine, once one becomes available.

“If supply of these components is disrupted you have to stop the production line,” Simchi-Levi says. “And many companies don't have a good enough understanding of their own supply chains to know who are the suppliers of their suppliers.”

Three stages

For small-molecule drugs such as remdesivir or hydroxychloroquine, production broadly involves three stages. The first yields the active ingredient in the drug; the second modifies the drug to make it stable and readily absorbed by the body; and the third packages the drugs, for example into tablets or vials. This takes place under the watchful eye of regulators, who periodically inspect facilities to ensure that quality and safety standards are maintained.

Relatively few sites are approved by regulators to make drugs, meaning that when one site fails an inspection – or when

“There's high dependency on only a few sites for manufacturing.”

more facilities are needed to crank out higher volumes of a particular drug – it can be difficult to find a replacement. “That can be pretty significant,” says Simchi-Levi. “There's high dependency on only a few sites for manufacturing.”

Production can be even more troublesome for more complex therapies, such as proteins or antibodies. Researchers are hopeful that antibodies that block certain immune-system processes will help against COVID-19, by restraining the out-of-control immune responses. Genentech in South San Francisco, California, makes one such antibody, called tocilizumab (Actemra), which blocks the activity of an immune-system regulator called IL-6. Tocilizumab is already approved for use against some forms of arthritis, but if it is found to be useful against COVID-19,

production would need to be vastly scaled up.

Antibody treatments such as tocilizumab are made in cells grown in culture, most often in Chinese hamster ovary cells. Antibodies are increasingly used to treat a range of diseases, from various forms of cancer to arthritis, and research has boosted production yields. About ten years ago, a manufacturer might expect to get less than 1 gram of antibody per litre of cell culture; now they typically extract 5 grams or more from the same volume, says Charles Christy, head of commercial solutions at the chemicals firm Lonza in Visp, Switzerland.

A 2,000-litre culture might produce enough antibody to fuel an early clinical trial, but drug-makers can scale up to as much as 20,000 litres of culture grown in giant steel vats to handle larger trials and commercialization.

Because antibody drugs are now such a large part of the pharmaceutical industry, there tend to be multiple suppliers for key reagents, Christy says. But you can always be blindsided, he says. “We and others are looking very hard at our supply chain.”

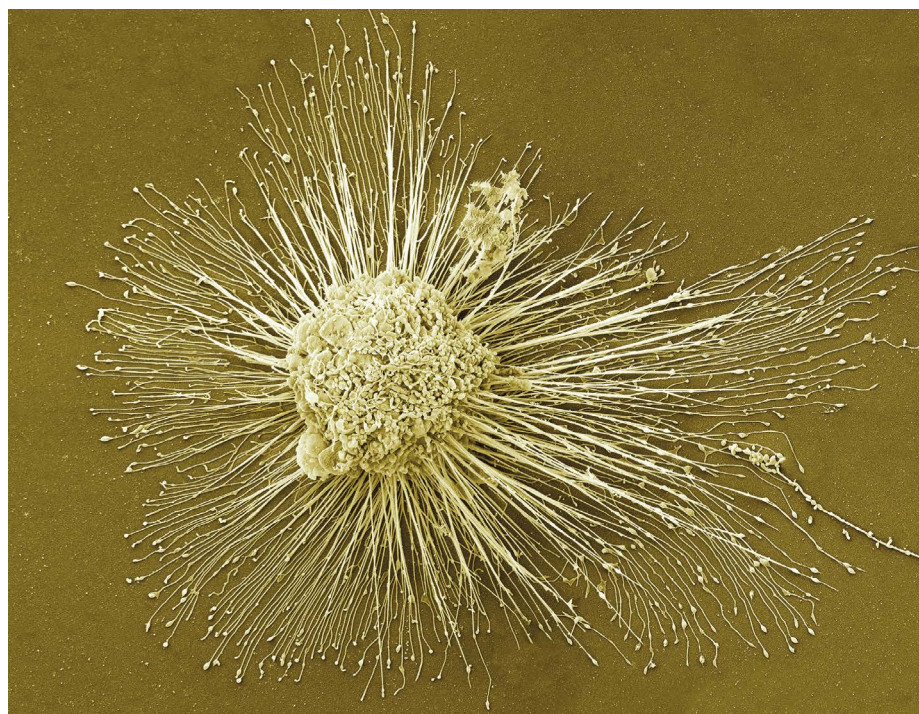
Tocilizumab has not yet been shown to help people with COVID-19, but Genentech says that it has already increased supply by 50% and is working to raise capacity further.

Massive demand

But even when companies work proactively to build supply, demand will almost certainly outstrip initial supplies of any compound found to be effective against COVID-19. That raises the spectre of determining who will be first to receive the treatments. Complaints have already surfaced about the allotment of remdesivir. Gilead has donated its stocks of the drug to treat COVID-19, with about 40% – enough to treat 78,000 people – going to the United States. The US government has been distributing those vials to individual states, but some hospitals have complained about lack of access.

Gilead also announced this week that it had entered into agreements with five makers of generic drugs. Those manufacturers can produce remdesivir for distribution in 127 countries that have limited access to health care, without paying royalties to Gilead. The agreement will remain in place until the global health emergency ends, or another treatment or a vaccine is found for COVID-19.

Concerns about access to pandemic medicines have arisen before, for example during the H1N1 influenza outbreak in 2009, says Emanuel, when countries raced to stockpile the influenza drug Tamiflu. “It was a free-for-all,” he says. Those issues have never been fully addressed because the outbreak ended quickly. “People move on and no one stays around long enough to solve the problem,” Emanuel says. “That will not happen here. We will be in this problem for a number of years.”



Heart muscle derived from induced pluripotent stem cells.

'REPROGRAMMED' STEM CELLS FOR HEART DISEASE TESTED IN CHINA

But there is no way to confirm that the unpublished trial using induced pluripotent stem cells works.

By Smriti Mallapaty

Two men in China were the first people in the world to receive an experimental treatment for heart disease based on 'reprogrammed' stem cells, and they have recovered successfully one year later, says the cardiac surgeon who performed the procedures. In May last year, the men were injected with heart muscle cells derived from induced pluripotent stem (iPS) cells, the surgeon told *Nature* – the first known clinical application of iPS-cell technology for treating damaged hearts.

No results have yet been published, so researchers not involved in the work have cautioned that there is no way to confirm whether the treatment works, including whether the reported benefits are due to the iPS-derived cells or simply to the heart bypass that accompanied the treatment.

But the surgeon, Wang Dongjin at Nanjing Drum Tower Hospital, spoke to *Nature* in detail about the procedure and about the patients' conditions. And one of the men – Han Dayong, a 55-year-old electrician from Yangzhou in

eastern China who received the treatment alongside a heart bypass – says he is very satisfied with the outcome. Before the surgery, Dayong remembers being tired and often out of breath. Now he can go for walks, climb stairs and sleep through the night. "It was beyond my expectations," he says.

The team behind the treatment plans to publish the results from the two recipients later this year, says Wang Jiaxian, who heads HELP Therapeutics, a biotechnology company based in Nanjing that supplied the heart muscle cells, known as cardiomyocytes, used in the study. The group also has approval to expand its study to include a further 20 patients, he says.

The trial in China is not the only one that is ongoing. In January, a cardiac surgeon in Japan, Yoshiki Sawa, introduced iPS-derived cardiomyocytes designed to treat heart disease into a patient. His team is using an alternative approach in which sheets of cells are grafted onto the heart rather than injected.

For decades, researchers have been trying to treat heart disease – a leading cause of death worldwide – using adult stem cells. They

hoped that the cells would morph into muscle cells once inserted into the heart.

But after trials in people proved inconclusive, researchers turned to iPS cells. These are created by inducing adult cells to revert to an embryonic-like state, from which they can develop into other cell types, such as cardiomyocytes.

Evidence in rodents and monkeys suggests that introducing iPS-cell-derived cardiomyocytes directly into the heart does regenerate muscle tissue and improve the organ's function¹. Researchers hope that the first trials in people will reveal the same.

"These are really exciting times," says Wolfram-Hubertus Zimmermann, a pharmacologist at the University Medical Centre Göttingen in Germany.

As well as the iPS-cell pilot study under way in Japan, several others are planned in France and the United States. Zimmermann is also planning one in Germany.

Safety first

That there is a trial ongoing in China came as a surprise to many, who didn't know that researchers there had overcome one of the field's biggest challenges – the need to produce large numbers of iPS-cell-derived cardiomyocytes that are pure enough to be used in people. This takes a lot of time and effort to get right, so very few companies or research groups have successfully done it, says Charles Murry, a pathologist at the University of Washington in Seattle who also plans to inject cells into people's hearts.

Wang Jiaxian says that his company has been developing the cells for almost four years.

Wang Dongjin told *Nature* that he injected some 100 million cardiomyocytes, derived from iPS cells created using cells donated by a healthy person, around the damaged heart tissue of his two patients. At the same time, both men, who had severe heart disease, underwent a coronary-artery bypass operation, in which vessels from elsewhere in the body are transplanted onto the artery to improve blood flow.

Wang says his goal was to assess the safety of the cell injections, and that he was encouraged when his patients' heart function improved significantly after the operations. Neither patient has developed tumours, he adds, which can be a risk of using pluripotent stem cells.

To prevent the body from attacking the cardiomyocytes, Wang says, both patients took immunosuppressant drugs. One took them for a month; the other had to stop after a week owing to side effects.

Wang also says that the procedure did not cause sustained dysfunction in heart rhythm. Zimmermann says that is a sign that it's safe, but it needs to be tested in more people.

Murry adds that the health benefits that the

patients have reported cannot be attributed to the reprogrammed cells alone, because the men also received a coronary bypass. “If you do two things to somebody and they get better, you can’t say which one caused it,” he says.

Researchers are divided on the best way to introduce cardiomyocytes into the heart. Injecting them is typically less intrusive than grafting sheets of cells, because it doesn’t require surgery, although the Chinese patients did have bypass operations. Proponents of injections also argue that in animals, the procedure has allowed the tissue to better integrate into the heart and produce new muscle².

But Philippe Menasché, a cardiac surgeon at the University of Paris, says that the injections puncture the heart in multiple locations,

which might damage the tissue.

In January, Sawa, a surgeon at Osaka University in Japan, trialled the alternative approach; grafting sheets of 100 million cardiac muscle cells onto a patient’s diseased heart. Sawa says the recipient moved out of intensive care within a few days. He plans to conduct the procedure in a further eight people.

Work in animals shows that more cells tend to survive being transplanted in sheets or patches than survive injection. But studies have also found that such grafted cells do not beat in synchrony with the heart³.

1. Liu, Y.-W. *et al. Nature Biotechnol.* **36**, 597–605 (2018).
2. Gerbin, K. A., Yang, X., Murry, C. E. & Coulombe, K. L. *PLoS ONE* **10**, e0131446 (2015).
3. Zimmermann, W.-H. *Curr. Opin. Physiol.* **14**, 70–77 (2020).

CORONAVIRUS BLOOD-CLOT MYSTERY INTENSIFIES

Research begins to pick apart the mechanisms behind a deadly COVID-19 complication.

By Cassandra Willyard

Purple rashes, swollen legs, clogged catheters and sudden death – blood clots, large and small, are a frequent complication of COVID-19, and researchers are just beginning to untangle why.

COVID-19’s impact on the respiratory system has gained much attention. But for weeks, reports have been pouring in of the disease’s effects throughout the body, many of which are caused by clots. “This is like a storm of blood clots,” says Behnood Bikdeli, a cardiologist at Columbia University in New York City. Anyone with a severe illness is at risk of developing clots, but people hospitalized with COVID-19 seem to be even more susceptible.

Scientists have a few plausible hypotheses to explain the phenomenon, and they are just beginning to launch studies aimed at gaining mechanistic insights. But with the death toll rising, they are scrambling to test clot-curbing medications.

Double whammy

Blood clots, jelly-like clumps of cells and proteins, are the body’s mechanism to stop bleeding. It’s not just their presence that has puzzled scientists: it’s how they show up. “There are so many things about the

presentations that are a little bit unusual,” says James O’Donnell, director of the Irish Centre for Vascular Biology at the Royal College of Surgeons in Dublin.

Blood thinners don’t reliably prevent clotting in people with COVID-19, and young people are dying of strokes caused by the blockages in the brain. And many people in hospital have drastically elevated levels of

“This is not what you’d expect to see in someone who just has a severe infection.”

a protein fragment called D-dimer, which is generated when a clot dissolves. High levels of D-dimer seem to be a powerful predictor of mortality in hospitalized people infected with coronavirus (L. Zhang *et al. J. Thromb. Haemost.* <https://doi.org/dv34>; 2020).

Researchers have also observed miniature clots in the body’s smallest vessels. “This is not what you’d expect to see in someone who just has a severe infection,” says Jeffrey Laurence, a haematologist at Weill Cornell Medicine in New York City. It’s a “double hit”, says O’Donnell. Pneumonia, which can be caused by COVID-19, clogs the tiny sacs in the lungs with fluid or pus, and microclots restrict

oxygenated blood from moving through them.

Why this clotting occurs is still a mystery. One possibility is that SARS-CoV-2, the coronavirus responsible for COVID-19, is directly attacking the endothelial cells that line the blood vessels. Endothelial cells harbour the same ACE2 receptor that the virus uses to enter lung cells. And there is evidence that endothelial cells can become infected: researchers at University Hospital Zurich in Switzerland and Brigham and Women’s Hospital in Boston, Massachusetts, observed SARS-CoV-2 in endothelial cells inside kidney tissue (Z. Varga *et al. Lancet* **395**, 1417–1418; 2020).

In healthy individuals, the blood vessel is “a very smoothly lined pipe”, says Peter Liu, chief scientific officer at the University of Ottawa Heart Institute in Canada. The lining actively stops clots from forming. But viral infection can damage these cells, prompting them to churn out proteins that trigger the process.

The virus’s effects on the immune system could also affect clotting. In some people, COVID-19 prompts immune cells to release a torrent of chemical signals that ramps up inflammation, which is linked to coagulation and clotting through a variety of pathways. And the virus seems to activate a defence mechanism that sparks clotting, known as the complement system. Laurence’s group found that small, clogged vessels in lung and skin tissue from people with COVID-19 were studded with proteins. All these systems – complement, inflammation, coagulation – are interrelated, says Agnes Lee, director of the Hematology Research Program at the University of British Columbia in Vancouver, Canada. “In some patients with COVID, all of those systems are kind of in hyperdrive.”

Race to new therapies

Even as researchers begin to unravel how clotting occurs in people with COVID-19, they’re sprinting to test new therapies aimed at preventing and breaking up clots.

At Columbia University, researchers are launching a clinical trial to compare the standard clot-preventing doses of blood thinners with a higher dose in people who are critically ill with COVID-19. Similar trials are planned for Canada and Switzerland. And scientists at Beth Israel Deaconess Medical Center in Boston have begun enrolment for a clinical trial to evaluate an even more powerful clot-busting medication called tissue plasminogen activator, or tPA.

Scientists hope that these trials and others will provide the data necessary to help physicians to make difficult treatment decisions. Lee worries about the amount of ‘reactionary medicine’ happening. “People are changing their therapeutic approach in reaction to their local and personal experience,” she says. She understands the impetus, “but we have to remember the main thing is first do no harm”.

ANTI-VACCINE MOVEMENT MIGHT UNDERMINE PANDEMIC EFFORTS

Studies of social networks show that opposition to vaccines is small but far-reaching – and growing.

By Philip Ball

As scientists work to create a vaccine against COVID-19, a small but fervent anti-vaccination movement is marshalling against it. Campaigners are seeding outlandish narratives: they falsely say that coronavirus vaccines will be used to implant microchips into people, for instance. In April, some carried placards with anti-vaccine slogans at rallies in California to protest against the state's lockdown. Last week, a now-deleted YouTube video promoting wild conspiracy theories about the pandemic and asserting (without evidence) that vaccines would "kill millions" received more than eight million views.

It's not known how many people would actually refuse a COVID-19 vaccine – and general support for vaccines remains high. But some researchers studying vaccine-opposition movements are concerned that the messages could undermine efforts to establish herd immunity to the new coronavirus. Online opposition to vaccines has rapidly pivoted to talk of the pandemic, says Neil Johnson, a physicist at George Washington University in Washington DC, who is studying the campaigners' tactics. "For a lot of these groups, it's all about COVID now," he says.

Groups opposing vaccines are small, but their online-communications strategy is worryingly effective and far-reaching, a report from Johnson's team suggests. Before the SARS-CoV-2 virus emerged, Johnson's team began mapping a network of views on vaccination, on Facebook. The researchers investigated more than 1,300 pages, followed by about 85 million individuals.

"For a lot of these groups, it's all about COVID now."

Their findings, published on 13 May, suggest that anti-vaccination pages tend to have fewer followers than pro-vaccination ones, but are more numerous (see 'Online competition between vaccine views'). They are also more often linked to from other Facebook pages – such as parent associations at schools – whose stance on vaccination is undecided (N. F. Johnson *et al. Nature* <http://doi.org/ggvvjx>; 2020).

By contrast, pages that explain the scientific case for vaccination are linked in a network that is largely disconnected from this "main battlefield" for public sentiment, as Johnson puts it. An extrapolation using computer simulations suggests that opposition to vaccines might dominate the network of views on the

subject within ten years, the team writes.

The work shows that "the pro-vaccine community are basically sticking to their narrative and talking to each other, and not reaching out and being responsive to the narratives that are out there among the undecided", says Heidi Larson, who directs the Vaccine Confidence Project, a group that monitors public trust in vaccines, at the London School of Hygiene and Tropical Medicine.

The issue isn't confined to Facebook. On 1 April, Johnson's team released a preprint of a study on online messaging about COVID-19 (N. Velásquez *et al.* Preprint at <https://arxiv.org/abs/2004.00673>; 2020). That report, which has not yet been peer-reviewed, suggests that, across different social-media platforms, links are growing between anti-vaccine groups debating COVID-19 and far-right extremists.

To counter anti-vaccine sentiment, scientists need an understanding of how the online map developed, says Bruce Gellin, president of global immunization at the Sabin Vaccine Institute in Washington DC. "We need to understand what it is about the conversations and content [around anti-vaccination] that compels people to listen and share it with others," he says.

Varied, emotive messages

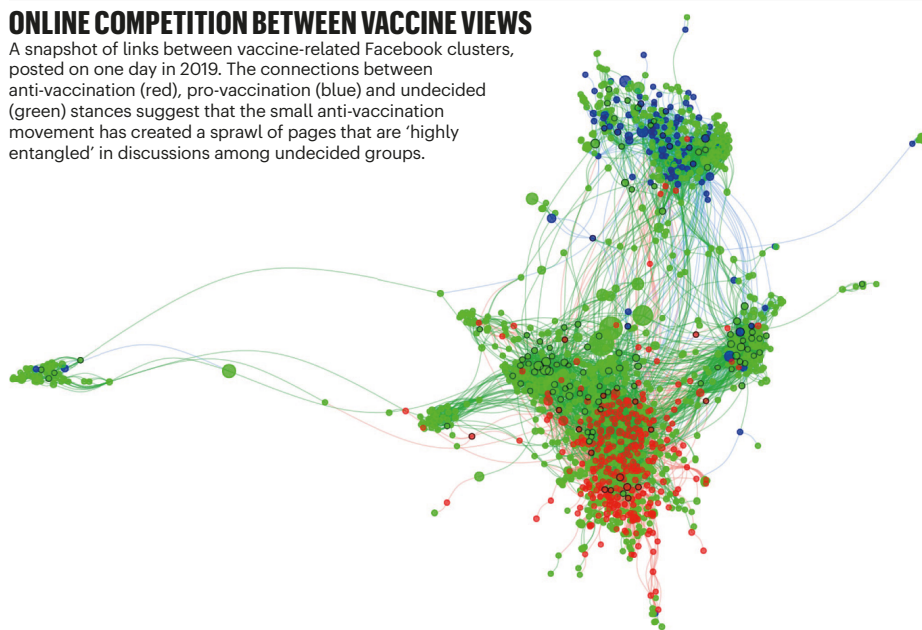
Pro-vaccine groups have a simple message: vaccines work and save lives. Anti-vaccine narratives are numerous – from sowing worries about children's health to advocating alternative medicines and linking immunizations to conspiracy theories. Anti-vaccine campaigners tend to win converts with personalized, emotive messages, says Larson; these are built not necessarily on fear (such as "Vaccines will kill you."), but on appeals to the heart ("Do you love your children?"). The public-health community, meanwhile, has simply been trying to get more people vaccinated, she says – which might lead to a feeling that they are just trying to get their numbers up. "The approach needs to be quite different with people who are undecided," she says. Vaccine-advocacy organizations are "not listening to concerns and questions".

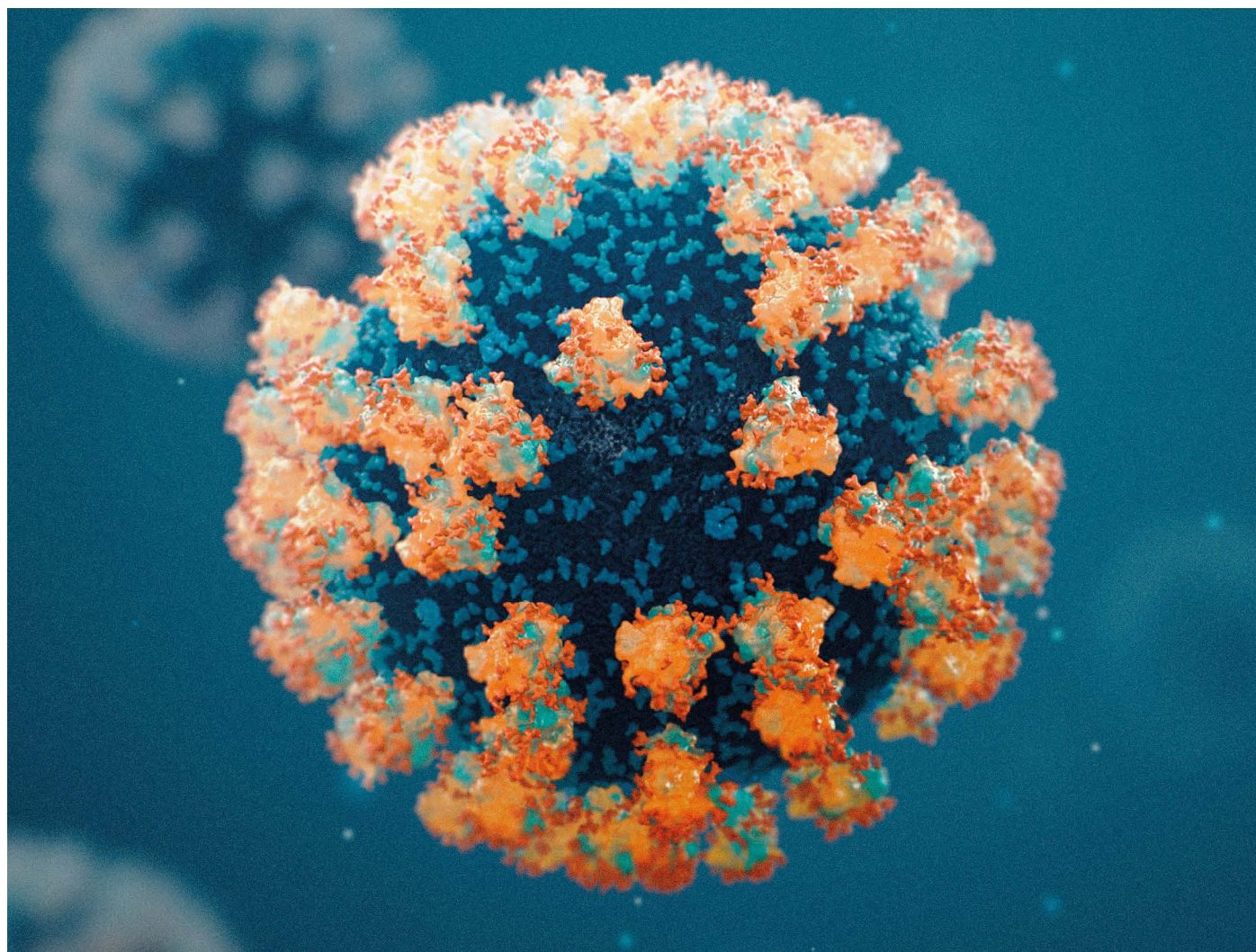
Overall, most people support vaccines, points out Gellin, and are likely to do so in this pandemic. Still, global vaccination rates have plateaued in the past two decades, Larson says. Both she and Gellin worry that another reason for public suspicion about a COVID-19 vaccine might be the speed of its development. "We should be very clear and transparent about the development process," says Gellin. "Otherwise, when it shows up, people will ask, 'How can we be sure no shortcuts were taken?'".

The messaging around a vaccine will also need to be carefully thought out. If there are already fewer COVID-19 infections by then, it's going to be a hard sell, says Larson. "The thing that's going to change people's minds is if the government says that if you have the vaccine, you can go to work," she says.

ONLINE COMPETITION BETWEEN VACCINE VIEWS

A snapshot of links between vaccine-related Facebook clusters, posted on one day in 2019. The connections between anti-vaccination (red), pro-vaccination (blue) and undecided (green) stances suggest that the small anti-vaccination movement has created a sprawl of pages that are 'highly entangled' in discussions among undecided groups.





COGNITION STUDIO INC.

CORONAVIRUS PIECE BY PIECE

Biologists are working at breakneck speed to solve the structures of key SARS-CoV-2 proteins and use them against the virus. **By Megan Scudellari**

Lying in bed on the night of 10 January, scrolling through news on his smartphone, Andrew Mesecar got an alert. He sat up. It was here. The complete genome of a coronavirus causing a cluster of pneumonia-like cases in Wuhan, China, had just been posted online.

Around the world, similar notifications appeared on the devices of scientists who first crossed swords with coronaviruses in the 2003 outbreak of SARS (severe acute respiratory syndrome) and then again with MERS (Middle East respiratory syndrome) in 2012. Instantly, the researchers mobilized against a new adversary. “We always knew that this was going to come back,” says Mesecar, head of biochemistry at Purdue University in

West Lafayette, Indiana. “It’s what history has shown us.”

In Lübeck, Germany, Rolf Hilgenfeld stopped packing boxes for his retirement and started preparing buffers for crystallography. In Minnesota, Fang Li stayed up all night analysing the new genome and drafting a manuscript. In Shanghai, China, Haitao Yang rallied a dozen graduate students to clear their schedules. In Texas, Jason McLellan instructed laboratory members to start assembling gene sequences from the viral genome.

Within 24 hours, a network of structural biologists around the world had redirected their labs towards a single goal – solving the protein structures of a deadly, rapidly spreading new contagion. To do so, they would need to sift through the 29,811 RNA bases in the

virus’s genome, seeking out the instructions for each of its estimated 25–29 proteins. With those instructions in hand, the scientists could recreate the proteins in the lab, visualize them and then, hopefully, identify drug compounds to block them or develop vaccines to incite the immune system against them.

“Here we go,” thought Mesecar. “I’d better get some sleep.”

11 January: 41 confirmed cases of COVID-19 worldwide

Mesecar woke at 6 a.m. the next day, turned on the coffee pot and began blasting through the new genome looking for recognizable protein sequences. It didn’t take long. He had spent 17 years studying coronaviruses, and the new virus’s genome looked very familiar.

“Holy shit,” he thought. “This is the same thing as SARS.”

Right away, Mesecar contacted Karla Satchell, a microbiologist at Northwestern University in Chicago, Illinois. Satchell is co-director of the Center for Structural Genomics of Infectious Diseases (CSGID), a consortium of eight institutions set up exactly for moments like this – to rapidly investigate the structures of emerging infectious agents.

To solve the 3D structure of a protein at high resolution, scientists first design a gene construct – a circle of DNA containing the instructions for the protein, together with regulatory sequences to control where and how it is expressed. They then insert the construct into living cells, often the bacterium *Escherichia coli*, using the cells’ own machinery to churn out the desired protein. Next, they purify the protein so that they can visualize its structure using either of two methods. One is X-ray crystallography, which involves growing tiny crystals of pure protein and revealing their internal structure by bombarding them with X-rays from a high-energy electron beam. The other is cryo-electron microscopy (cryo-EM), a process of scanning flash-frozen proteins using a high-powered electron microscope.

Either process can take months, even years, for an unfamiliar protein. Luckily, many of the new coronavirus proteins were familiar, with 70–80% sequence similarity to SARS-CoV, the virus that caused the 2003 SARS outbreak. By 7:30 a.m., Mesecar and his team had begun designing gene constructs for the new viral proteins, and even predicted which of their existing coronavirus inhibitors might block these proteins.

Satchell, who had been following early news reports about the virus, organized a virtual meeting of consortium members to start solving the virus’s proteins. “We’ve thrown the weight of every investigator at every site behind COVID,” says Satchell. Mesecar, a CSGID investigator, started with M^{pro}, the virus’s main protease, an enzyme that cuts out proteins from a long strand that the virus produces when it invades a cell, like a tailor cutting out pattern pieces. Without M^{pro}, there is no viral replication. Humans do not have a similar protease, so drugs targeting this protein are less likely to cause side effects.

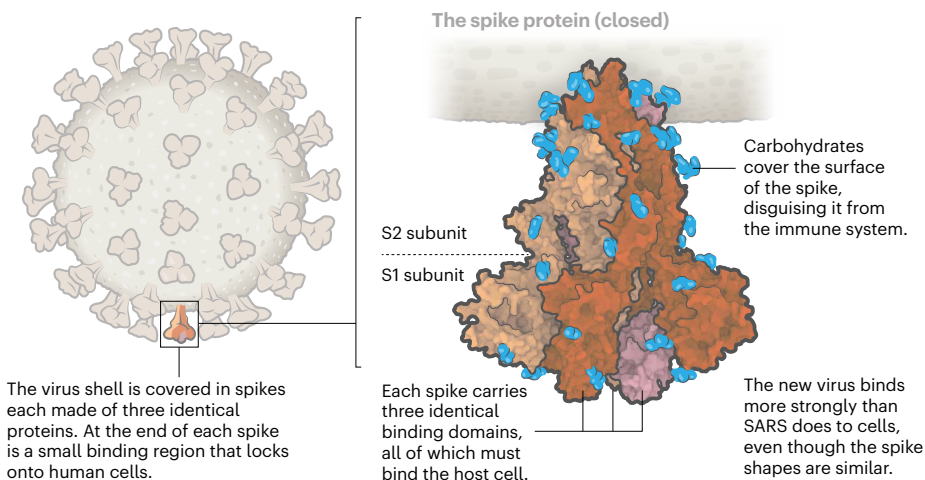
13 January: 42 confirmed cases

In McLellan’s molecular biosciences lab at the University of Texas at Austin, graduate student Daniel Wrapp spent the weekend designing a gene construct for another key protein – the outer, three-pronged spike that gives the coronavirus its crown-like appearance and name. Wrapp placed an order for the constructs with a commercial firm that Monday, 13 January.

McLellan had been involved in determining

THE KEY CORONAVIRUS PROTEINS

Researchers are racing to visualize and understand the proteins used by SARS-CoV-2 to enter cells and replicate. That information could be crucial for making drugs and vaccines to stop the virus.



the structures of two other coronavirus spikes – from HKU1, a cause of common colds¹, and from the MERS virus². The work was done in collaboration with structural biologist Andrew Ward at the Scripps Research Institute in La Jolla, California, and virologist Barney Graham at the US National Institute of Allergy and Infectious Diseases’ Vaccine Research Center in Bethesda, Maryland. So, the group knew how to tweak the spike protein’s genetic sequence so that it would stabilize in a pre-fusion shape – the form it adopts before it docks onto a host cell. “Our ability to get this particular structure was based upon all our prior knowledge from working on HKU1 and MERS and SARS,” says McLellan.

While McLellan’s team waited for the construct to arrive, Graham called Moderna Therapeutics, a drug-discovery company in Cambridge, Massachusetts, with which the Vaccine Research Center had been working on a pandemic-preparedness project. On 13 January – before any spike protein had been made – Moderna began preparing its manufacturing facilities to make a coronavirus vaccine based on that protein.

26 January: 2,014 confirmed cases

At ShanghaiTech University in China, Zihao Rao, Haitao Yang and their colleagues worked day and night, sacrificing their week-long Chinese Lunar New Year holiday, to solve the M^{pro} structure and those of another trio of proteins that the coronavirus uses to replicate.

Using X-ray data acquired at the Shanghai Synchrotron Radiation Facility and the National Center for Protein Science Shanghai – which both allocated special beam time for the project – the team solved the crystal structure of M^{pro} bound to an inhibitor³. In 2003, it had taken them two months to solve the structure of the SARS-CoV main protease. This time, it took one week.

M^{pro} in coronaviruses is made up of two identical subunits and looks like a moth-eaten heart, with an active enzyme site on each side of the structure. On 26 January, Rao and Yang submitted the M^{pro} structural data to the Protein Data Bank (PDB), an open-access digital resource for 3D structures of biological molecules. By 5 February, the data had been processed and the final structure was released online – not a moment too soon, says Yang. The laboratory had already received an overwhelming 300 requests for the structure.

While working on M^{pro}, Rao contacted a former co-worker, David Stuart, a structural biologist at the University of Oxford, UK, who is life-sciences director at Diamond Light Source, the United Kingdom’s synchrotron facility. The UK and Shanghai groups began collaborating closely to share advice and avoid overlap, says Martin Walsh, deputy life-sciences director at Diamond. “We keep each other up to date on things, and try to benefit from the different approaches they’re using and we’re using.”

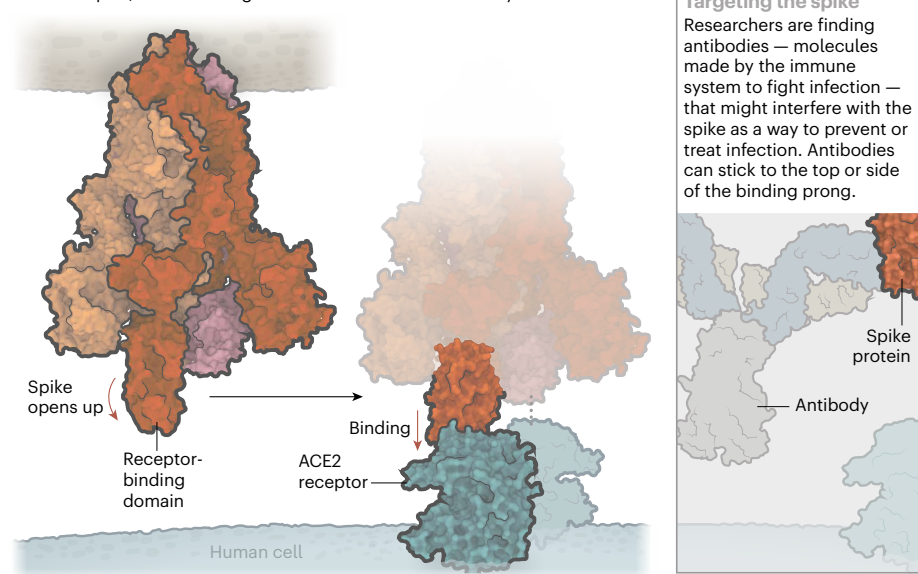
Because the Shanghai team solved M^{pro} in complex with an inhibitor, the Diamond team decided to focus on crystallizing the protein with no molecule attached, hoping to identify active sites to which potential drug compounds might bind. Over two weeks, Walsh’s group ran 17,000 experiments to hit on the best recipe for precipitating the unbound protein into a crystal.

1 February: 11,953 confirmed cases

In Hilgenfeld’s lab at the University of Lübeck, researcher Linlin Zhang had taken to phoning the company making the M^{pro} gene construct daily until it finally arrived. Thanks to the lab’s experience crystallizing other coronavirus proteases, Zhang grew M^{pro} crystals in 10 days, and on 1 February, she took the precious samples to the BESSY II synchrotron in Berlin, which

THE SPIKE LOCKS ON

Binding regions on the tip of the spike open out to attach to the human ACE2 receptor, found on lung cells and elsewhere in the body.



opened up a beamline especially for the project.

In addition to focusing on the unbound M^{pro} structure, Hilgenfeld docked a small-molecule inhibitor called 13a, which he had designed to inhibit the MERS virus, into the protein's active site. It wasn't a perfect fit, so the team altered a residue on the compound and named it 13b. This one "fit nicely", says Hilgenfeld, and in ten more days his team had solved the structure of M^{pro} bound to the inhibitor⁴.

McLellan's group in Texas was solving the spike protein structure at similar speed. As soon as the group had finished gathering high-resolution electron-microscopy data of the stabilized spike, thanks to a multimillion-dollar cryo-EM facility at the university, McLellan sent the data to Graham at the Vaccine Research Center.

Vaccines are often based on presenting parts of a virus to the human immune system to provoke a response, and the spike protein is an obvious candidate because it has a crucial role in infection.

The spike is formed of three identical molecules stuck together in the shape of a pyramid, with a hinge-like trapdoor. This opens to expose a portion that grabs onto a receptor on a human cell (see 'The key coronavirus proteins'). Graham and McLellan's past work on a similar protein⁵ suggested that presenting the spike protein in its pre-grab state would provoke the human immune system. From the complete structure, Graham could see that McLellan's gene construct made a high-quality protein arranged in the right conformation. "It was really, really important to have that electron-microscopy information," says Graham.

Graham tested the spike protein in mice, working to improve its expression levels and the strength of its effect on the immune

system, and sent the sequence to Moderna, where the production line was ready and waiting. On 7 February, Moderna completed its first batch of the vaccine based on that protein.

Meanwhile, on 10 February, just 12 days after harvesting the protein, McLellan and his group submitted its cryo-EM structure⁶ to the PDB. By studying the spike in detail, they found that it binds to its human cell receptor, a protein called ACE2, at least ten times more tightly than SARS-CoV does.

At the University of Minnesota in Saint Paul, Li's team was on its way to working out why. On 11 February, Li and his colleagues began collecting X-ray data from the spike protein using the Advanced Photon Source (APS), the synchrotron facility at the US Department of Energy's Argonne National Laboratory near Chicago, Illinois. By 13 February, the researchers had defined the small, important spot where the spike protein locks on to the ACE2 receptor⁷ (see 'The spike locks on'). They found that the new coronavirus spike protein has small molecular differences in its binding region compared with that of SARS-CoV, which might be why the new virus attaches to ACE2 more strongly. These changes could also explain why it seems to infect cells better and spreads faster than the SARS virus. That same week, the virus also got a name: SARS-CoV-2.

18 February: 73,332 confirmed cases

By mid-February, protein structures were pouring out. On 18 February, Hilgenfeld, Zhang and their colleagues submitted a paper⁴ on the M^{pro} structure alone and bound to 13b, and posted the preprint on the bioRxiv server on 20 February. "It was pretty fast," Hilgenfeld admits. "The longest time period was just getting it published." That same day, the Diamond team released the high-resolution

crystal structure of unbound M^{pro} on its website (go.nature.com/2z41uwj).

To support US teams, the APS and other national synchrotrons coordinated their schedules to ensure there would be no interruption in beamtime if one facility had to close for maintenance or because of a local outbreak. "Our goal is just to keep the research going," says Stephen Streiffer, director of the APS. "The rate at which people are working at this is an order of magnitude faster than they've been able to work on other problems."

So far, the CSGID consortium has solved 12 unique SARS-CoV-2 protein structures, which are kept in a new online database with their accompanying genomic information (see <https://coronavirus3d.org>). "We've been part of projects like this on cancer, but it took five years to set that all up," says Adam Godzik, a bioinformatician at the University of California, Riverside, and a CSGID investigator. "This happened spontaneously in the course of months."

16 March: 167,515 confirmed cases

With 3D structures in hand, structural-biology teams moved straight to next steps. "Structures aren't everything," says Mesecar. "You want to get to compounds — to antivirals and vaccines."

On 16 March, just 65 days after the viral genome was released, clinicians gave the first dose of Moderna's vaccine candidate to a patient in a clinical trial funded by the US National Institutes of Health.

"It was a lot faster than even the fastest one we'd previously done," says Graham. Because of research on SARS and MERS, coronaviruses were probably the only viral family for which that was possible, he adds. "If it was a bunyavirus or an arenavirus, we would have been lost for two to three years."

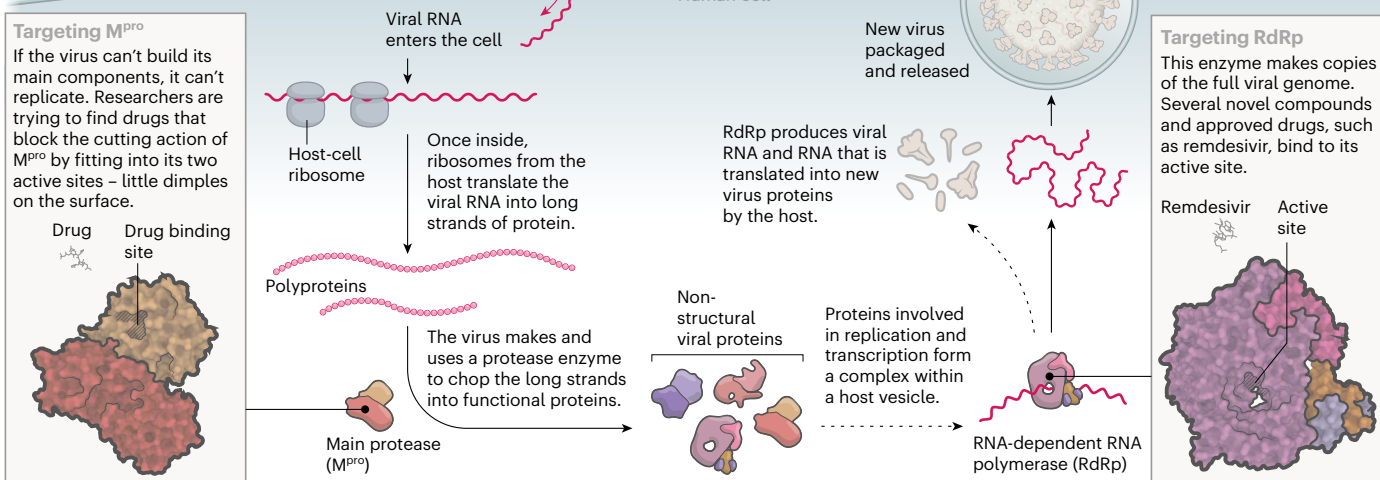
But even a vaccine developed at record-breaking speed is likely to be a slower solution than repurposing an approved drug, or at least finding one for which safety testing has begun. "That's absolutely going to be the fastest way to help patients sick in the hospital today," says Satchell.

That was exactly what Andrew Hopkins was planning. On 19 March, Hopkins, the chief executive of Exscientia, an artificial-intelligence drug-discovery company in Oxford, UK, took delivery of a large styrofoam cooler packed with dry ice. Inside was a library of 12,000 drug compounds known to be safe and ready for human use, sent from Scripps Research in California. The Exscientia team, working closely with Diamond, immediately began screening the collection against four of Diamond's structures: M^{pro} , the spike protein, a second protease and the replication-machinery complex. Exscientia is currently preparing to test compounds that bind to the first two proteins for antiviral activity, says Hopkins.

SOURCES: OPEN SPIKE: REF. 6; ACE2 BINDING: REF. 7; ANTIBODY BINDING: M. YUAN ET AL. SCIENCE 368, 630–633 (2020).

BREAKING THE CYCLE

Once the virus has fused with a host cell, the virus injects genetic material and uses the host machinery to make copies of itself. Many teams are studying viral proteins involved in replication.



Similarly, the ShanghaiTech team conducted virtual and high-throughput screening of a library of more than 10,000 approved drugs and compounds already in clinical trials, to see whether any would disable M^{pro}. They identified six promising candidates³. One of them, ebselen, is already in clinical trials for the treatment of bipolar disorder and hearing loss, and the group is preparing animal tests to study its activity *in vivo*, says Yang.

On 10 April, Rao, Yang and their collaborators published⁸ the structure of the virus's replication complex – a large protein called RNA-dependent RNA polymerase (RdRp, or nsp12) that forms a complex with two others, nsp7 and nsp8. They also modelled how it binds to the antiviral drug remdesivir, originally developed to treat Ebola and now in phase III trials for coronavirus. Another recently completed structure of the protein in complex with the drug⁹ could provide a template to help model and modify other existing antivirals.

22 April: 2,471,136 confirmed cases

The hard-core biochemistry of designing brand-new, custom drugs to inhibit SARS-CoV-2 proteins will take months, even years, but could eventually lead to the best-performing drugs against the infection (see 'Breaking the cycle').

The ShanghaiTech team and collaborators have designed and synthesized a series of compounds targeting the active site of M^{pro}. On 22 April, after much chemical tweaking, they published details of one that inhibits viral replication in cells and was not toxic when tested in rats and dogs¹⁰. The team will continue developing that compound as a drug candidate, says Yang.

The Diamond team has identified

91 chemical fragments – bits of molecules that are less than one-third the size of a normal drug – that bind to M^{pro}. Those fragments inspired the launch of a non-profit crowd-sourced initiative, the COVID Moonshot, to engage chemists around the world to use the fragments to design antiviral drug candidates. The initiative has received more than 4,600 design submissions, and several therapeutic possibilities are already emerging.

In Germany, researcher Katharina Rox at the Helmholtz Centre for Infection Research in Braunschweig tested Hilgenfeld's 13b

"Structures aren't everything. You want to get to compounds – to antivirals and vaccines."

compound in mice, showing that it was safe and accumulated well in the lungs⁴, a key infection site. Meanwhile, a compound that Mesecar developed to inhibit SARS-CoV, compound 77, has been shown in unpublished work to have antiviral activity against SARS-CoV-2 in cells, and he hopes to complete animal studies by the end of the summer.

14 May: 4,248,389 confirmed cases

Structural biologists continue to plug away at the remaining unsolved proteins in the coronavirus genome. These include ORF8, a protein whose function remains mysterious. "We predict it should be crystallizable, but nobody has done it, so we're trying," says Godzik.

In the United Kingdom, the Diamond team is screening various compounds against a second coronavirus protease. In Texas, McLellan has

shipped spike constructs to more than 100 labs worldwide. Many are looking for treatments, using the protein to fish antibodies out of the blood of people who have had COVID-19, and McLellan's team is now characterizing the first of these potentially therapeutic antibodies.

Hilgenfeld, who was officially scheduled to retire on 1 April as a result of a mandatory retirement policy, has packed up his office but continues to work. "I've been working on coronaviruses for 20 years, and most of the time it was neglected and not taken seriously," he says. "Now that it's happened, how can I leave?" His team is investigating other SARS-CoV-2 structures, including nsp3, a large protein that the virus uses to shut down host-cell defences.

The race against the virus can't afford to slow down anytime soon. As soon as countries start lifting restrictions on people's movement, the virus will return and "flip around the world again", says Satchell. "When that happens, it would be really great to have beautiful drugs that were designed specifically to target this coronavirus," she says. "But we need to do it fast."

Megan Scudellari is a science journalist in Boston, Massachusetts.

1. Kirchdoerfer, R. N. et al. *Nature* **531**, 118–121 (2016).
2. Pallesen, J. et al. *Proc. Natl Acad. Sci. USA* **114**, E7348–E7357 (2017).
3. Jin, Z. et al. *Nature* <https://doi.org/10.1038/s41586-020-2223-y> (2020).
4. Zhang, L. et al. *Science* **368**, 409–412 (2020).
5. McLellan, J. S. et al. *Science* **342**, 592–598 (2013).
6. Wrapp, D. et al. *Science* **367**, 1260–1263 (2020).
7. Shang, J. et al. *Nature* **581**, 221–224 (2020).
8. Gao, Y. et al. *Science* **368**, 779–782 (2020).
9. Yin, W. et al. *Science* <https://doi.org/10.1126/science.abd1560> (2020).
10. Dai, W. et al. *Science* <https://doi.org/10.1126/science.abb4489> (2020).



NURUL HIDAYAT/ANTARA FOTO/BISNIS INDONESIA VIA REUTERS

When Anak Krakatau in Indonesia erupted on 22 December 2018, part of the island collapsed into the ocean, causing a deadly tsunami.

THE VOLCANOLOGY REVOLUTION

Forty years after the Mount St Helens eruption galvanized volcano science, researchers are harnessing powerful new tools to forecast and understand eruptions. **By Jane Palmer**

Early in 2018, the volcano Anak Krakatau in Indonesia started falling apart. It was a subtle transformation – one that nobody noticed at the time. The southern and southwestern flanks of the volcano were slipping towards the ocean at a rate of about 4 millimetres per month, a shift so small that researchers only saw it after the fact as they combed through satellite radar data. By June, though,

the mountain began showing obvious signs of unrest. It spewed fiery ash and rocks into the sky in a series of small eruptions. And it was heating up. Another satellite instrument recorded thermal emissions from Anak Krakatau that reached 146 megawatts – more than 100 times the normal value. With the increased activity, the slippage jumped to 10 millimetres per month.

Then, on 22 December, the southern flank crashed into the sea, triggering a tsunami that

killed at least 430 people along the nearby coasts of Java and Sumatra. Although nobody foresaw that disaster, a 2019 study found that satellite and ground-based instruments had picked up a suite of precursory signals that could help forecast similar events in the future at Anak Krakatau and other peaks.

The unexpected collapse at Anak Krakatau shows some of the challenges facing researchers as they try to monitor thousands of

potentially dangerous volcanoes around the world – each one unique. But it also highlights several advances in the field that promise to give scientists a much better chance of forecasting disasters.

Volcanologists are making substantial headway, thanks to a torrent of data from satellites that can detect subtle movements of mountains, ground-based sensors that track molten rock moving deep underground, and gas-sniffing devices that drones can carry over seething mountains. And the theoretical understanding of volcanoes has grown markedly as researchers have learnt to combine all these data into models of what is happening within volcanic systems. Researchers are now experimenting with machine learning to sift through the flood of data to identify subtle patterns, such as the early movement of Anak Krakatau months before it showed signs of waking.

The field has made huge strides since the greatest volcanic crisis in US history exactly 40 years ago – the eruption of Mount St Helens on 18 May 1980 in Washington state. That event – which started with the largest landslide in recorded history – killed 57 people and blanketed much of Washington and nearby states with ash, shutting down the region for days. But it was also a turning point for volcanic science, sparking a huge influx of money and people into the field and setting the stage for rapid improvements in understanding.

Scientists had flocked to the mountain in the months before the blast and had carefully tracked its behaviour, including frequent earthquakes, gases fuming from its crater and an ominous bulge that swelled from its northern flank. “It was the first really significant eruption that was captured by modern-day scientific instrumentation,” says Seth Moran, scientist-in-charge at the US Geological Survey (USGS) Cascades Volcano Observatory in Vancouver, Washington. “And so, in a lot of ways, it’s become a benchmark for the ways that people go about looking at volcanoes around the world.”

The proliferation of ground- and space-based monitoring data since then, coupled with increases in computing power, has revolutionized scientists’ understanding of volcanic systems. Ultimately, researchers are hoping that new tools and techniques will nudge them closer to being able to assign probabilities to the chances of a volcano erupting in a given

time frame, much as meteorologists dole out the chances of rain or snow on any specific day.

“I think that when people look back on this period, they will imagine this is the golden era of physical volcanology,” says volcanologist Christopher Kilburn at University College London.

Historic blast

The first hints of trouble at Mount St Helens came on 16 March 1980, with a series of small earthquakes. Then, a week later, steam explosions burst through the ice on top of the volcano, carving out a crater that grew to 400 metres across within days. Teams of researchers arrived from the USGS and other institutions to keep vigil over the mountain. Planes flew

“Tall volcanoes collapse. They are not just growing, they’re collapsing.”

over the smoking crater to measure the gases escaping from the volcano, and seismometers registered the tremors from magma – molten rock – moving beneath the surface. Volcanologists climbed the mountain’s slopes to measure the bulging northern flank using tape measures and laser-surveying equipment.

Magma was clearly rising high in the volcano and pushing against the slope, and researchers warned that a major eruption could happen soon. But what happened next caught scientists by surprise.

At 8:32 a.m. on 18 May, a massive landslide crashed down the mountainside, taking the summit and snow and ice with it. The release in pressure uncorked the volcano, triggering a powerful explosion. A blast of rocks, ash, gas and steam was propelled upwards and outwards at supersonic speeds, and travelled as far as 25 kilometres northwards.

“We learnt from the May 18th eruption how unstable steep-sided volcanoes are, and how they can fail and generate a big surge or lateral blast,” says Don Swanson, a research geologist at the USGS Hawaiian Volcano Observatory, who was involved in monitoring the 1980 eruption. “What seems so obvious now, wasn’t obvious before that time.”

After the eruption, scientists analysed the landscape and found it littered with

hummocks – large hills and mounds that had been transported downslope in intact blocks. These features matched those found near many volcanoes around the world. And from the historical record, volcanologists recognized that around 1,000 similar landslides had taken place on more than 550 volcanoes. “Tall volcanoes collapse, they’re not just growing, they’re collapsing,” says volcanologist Thomas Walter at the German Research Centre for Geosciences in Potsdam.

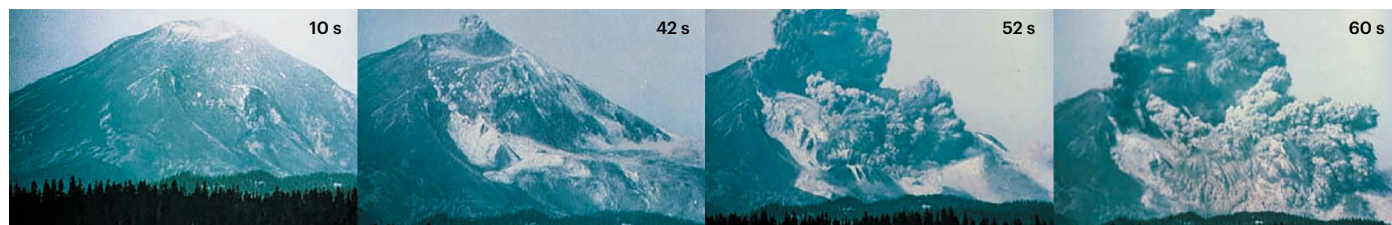
The eruption of Mount St Helens taught other lessons, such as the deadly impact of superheated volcanic ash and gas racing down the mountain at hurricane speeds, and the power of mudslides that destroyed everything in their path. The eruption also spurred a huge growth in volcanology. In the decade after the blast, the USGS established volcano observatories in the Pacific Northwest, Hawaii and Alaska.

Funding for the USGS’s volcanic hazards programme today is nearly ten times what it was before the Mount St Helens blast. And after a volcanic mudslide in Colombia killed 23,000 people in 1985, the USGS established the Volcano Disaster Assistance Program to help other countries prepare for volcanic crises – a project that soon proved its worth when USGS researchers worked with scientists in the Philippines in 1991 to assess the risk from Mount Pinatubo. Tens of thousands of people were evacuated from the region before the volcano’s cataclysmic eruption.

Researchers today rely on many of the lessons learnt at St Helens, Pinatubo and dozens of other volcanoes. Typically, seismic shaking is the first sign that a volcano is stirring. Eruptions occur when magma pushes to the surface, but even as magma begins to rise from Earth’s mantle, it can trigger quakes. Today, seismic networks are monitoring dozens of some of the most dangerous volcanoes around the world.

That same magma movement can cause volcanoes to inflate, as Mount St Helens did before its blast. Researchers can now record movements safely and continuously, using GPS receivers and, more recently, satellite-borne radar – which detected the movement at Anak Krakatau.

Even before warning signs can be seen or felt, rising levels of carbon dioxide from a volcano’s crater or vents can hint at trouble ahead. Magma contains dissolved gases and as this molten material rises and the pressure



On 18 May 1980, a giant landslide – the largest in recorded history – carried away the north flank of Mount St Helens, triggering an eruption.

Feature

decreases, gases separate off and travel upwards. Carbon dioxide, one of the least soluble of the volcanic gases, escapes first, while the magma is still deep in the volcano. “In principle, you should get a gas signal long before the magma reaches the surface in an eruption,” says volcanic-gas geochemist Alessandro Aiuppa at the University of Palermo in Sicily, Italy.

Historically, scientists had to collect gas samples from near the crater or vents – a dangerous task that yielded only episodic bites of information. Then, in 2005, Italian researchers designed an instrument – a multicomponent gas analyser system (Multi-GAS) – that is not much bigger than a shoebox. Volcanologists install these sensors near vents, and also mount them on drones that fly over active craters to measure the levels of five key gases emitted by volcanoes. “This has been a real revolution for volcanic-gas science because it means you can have a measurement of volcanic-gas composition every second, in real time, on your computer,” Aiuppa says.

Blast forecast

The Multi-GAS instruments had their trial by fire on Stromboli, a volcano off the north coast of Sicily. Italian scientists installed these sensors, along with cameras and spectrometers, on the volcano in 2005 and have collected gas data ever since. In February 2007, lava began to ooze out of the volcano in an effusive eruption. The researchers saw that carbon dioxide levels rose tenfold over the two weeks before the volcano erupted explosively on 15 March¹.

The findings allowed volcanologists to build a conceptual model of this complex volcano, in which explosions emanate from a deep magma chamber 7–10 kilometres below the summit. The researchers determined that the chances of an explosive eruption increase when carbon dioxide emissions top 2,000 tonnes per day.

In August 2019, Stromboli oozed lava again, and for the next two weeks the Italians tracked a slow, progressive increase in carbon dioxide. “So, we knew that something was going to happen,” Aiuppa says. The team increased its vigilance and also closely monitored ground-level changes using tiltmeters that measure subtle changes in ground angle. Eventually, what they saw made them certain that an explosion was coming soon, and they alerted the local authorities minutes before a blast on 28 August.

At Mount Etna on the Sicilian mainland, Italian researchers are tracking low-frequency sound waves – infrasound waves – that some volcanoes emit before they erupt. Scientists installed the system on Etna in 2008 and analysed its performance for 59 eruptions in the following 8 years. It successfully predicted 57 of the events, and sent messages to the researchers about an hour before each eruption². Given this success, in 2015 the team

programmed the system to send automatic e-mail and text-message alerts to the civil-protection department in Rome and to the city of Catania close to the volcano.

The researchers’ original motivation for developing the system was to find a way to detect eruptions at unmonitored volcanoes, because even remote blasts can have far-reaching impacts. The eruption of Eyjafjallajökull in Iceland in 2010 created an ash plume that disrupted air traffic across Europe for weeks. “Volcanic risk has no borders,” says Maurizio Ripepe, a geophysicist at the University of Florence, Italy, who helped to create the automated early-warning system on Etna.

Currently, fewer than half of the world’s active volcanoes on land have any sort of ground instrumentation, and in many cases this consists of just a few seismometers. But in the past decade, researchers have gained new ways to monitor all volcanoes using instruments mounted on satellites.

Data deluge

On 10 April 2020, Indonesia’s Anak Krakatau spewed a column of ash 500 metres into the sky and the Center for Volcanology and Geological Hazard Mitigation in Indonesia issued a level-2 alert, which signifies that the volcano has the potential to erupt but poses limited hazards.

After the deadly tsunami in 2018, German volcanologists had found a striking pattern in data recorded by the moderate-resolution imaging spectroradiometer (MODIS) on a NASA satellite. Infrared channels revealed that thermal emissions jumped in June 2018 (ref. 3). “The whole volcano was hot, the most intense

activity ever recorded,” says Walter. “So, this was clearly anomalous behaviour.”

The researchers also used satellite radar observations, which can detect small changes in vertical and horizontal motion, to find that the volcano’s flank was already slipping at a rate of 10 millimetres per month before it collapsed (see ‘Island on the move’).

The research demonstrated how, even when ground instrumentation is limited, scientists can learn about the lead-up to an eruption or volcanic landslide from satellites. “As volcanologists, we always used to say that we were data poor,” says Michael Poland, scientist-in-charge at the USGS Yellowstone Volcano Observatory in Vancouver, Washington. “But now the satellite data really expand our ability to see what volcanoes are doing.”

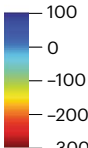
Volcanology got a huge boost in 2014 and 2016 with the launch of the European Space Agency’s Sentinel 1A and 1B radar satellites. Using the technique of interferometric synthetic-aperture radar, they can track movements of volcanoes at unprecedented resolution levels and at frequent time intervals (see ‘Inflation watch’). “These satellites can detect subcentimetre deformation of ground surfaces, meaning that we can see when the volcano is swelling,” says volcanologist Charles Mandeville, programme coordinator of the USGS Volcano Hazards Program. “There is a whole fire hose of such data being collected now.”

Researchers have combined radar data with satellite observations that record temperature and sulfur dioxide emissions to capture a multidimensional picture of what happens at volcanoes before and during eruptions. A

ISLAND ON THE MOVE

Satellite radar data reveal how the ground surface of Anak Krakatau, a volcanic island, shifted in the 12 months before an eruption on 22 December 2018. The southwestern region collapsed during the eruption, triggering a deadly tsunami. Such observations could help to forecast when a volcano will erupt.

Vertical movement
(mm yr⁻¹)



Movement
East or West
(mm yr⁻¹)

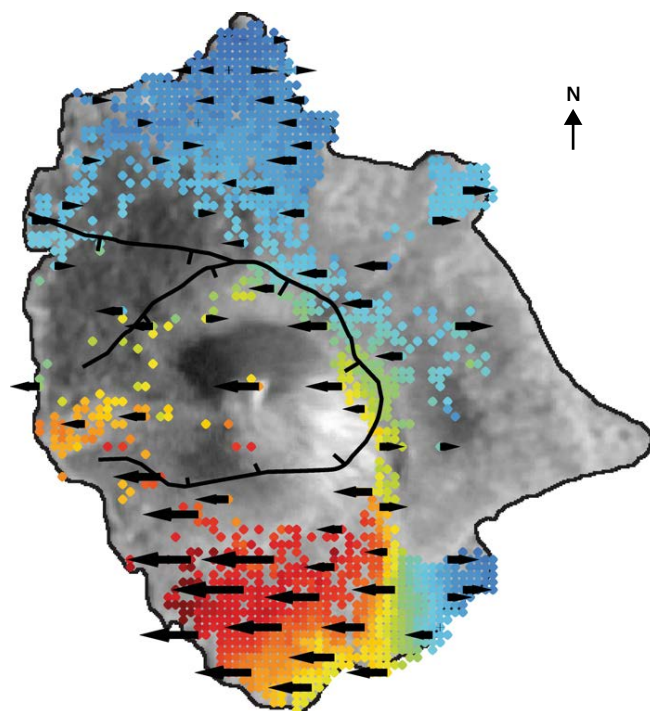

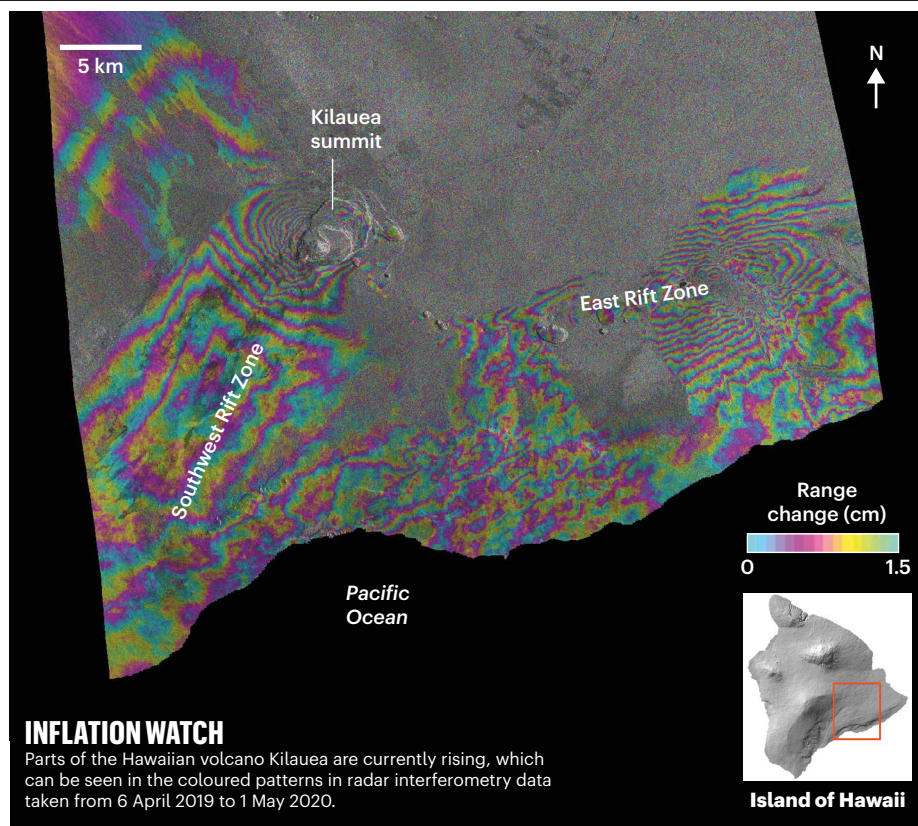


FIG. 3B WALTER, T. R ET AL. NATURE COMMUN. 10, 4339 (2019)



study of the 47 most active volcanoes in South America, which used 17 years of satellite data, showed that changes in at least one of these variables, and sometimes in all three, precede an eruption, sometimes years in advance⁴.

To exploit these data, many of which are freely available, Walter and colleagues have created a volcano-monitoring platform called MOUNTS (monitoring unrest from space). The platform uses data from the current suite of Sentinel satellites and ground-based earthquake information, and currently monitors 17 volcanoes, including Anak Krakatau.

As they started on the project, however, the researchers faced a new and unusual problem — too much data. The satellites provide torrents of readings, more than researchers can analyse using conventional methods. “There are so many volcanoes and so much data that we needed smarter ways of dealing with the data set,” Walter says.

In response to this challenge, researchers have turned to machine-learning techniques, a form of artificial intelligence in which computer algorithms such as neural networks can be trained to pick out patterns in data. Juliet Biggs, a volcanologist at the University of Bristol, UK, and her colleagues have created a neural network that has churned through some 30,000 Sentinel-1 images of more than 900 volcanoes and flagged about 100 images as needing more attention. Of those images, 39 showed real ground distortions⁵, meaning that the AI system had reduced the workload for the volcanologists by a factor of nearly 10. Now, they are testing their system on some

half a million images from more than 1,000 volcanoes.

“You just can’t look at every image,” Poland says. “I see machine learning as having a real impact in filtering through these massive volumes of data.”

For the MOUNTS platform, scientists have also developed a neural network to search for large shape changes. Other groups are trying to develop algorithms that can sift through temperature or gas-emission data from satellites.

When Anak Krakatau sprang back into action on 10 April this year, Walter was quick to monitor the situation remotely by analysing the satellite data. Because visibility was low, he had to rely on the radar data, which can penetrate thick clouds. The information will help scientists understand the behaviour of Anak Krakatau and in the future it might be used to help create a tsunami early-warning system for landslides from the Indonesian volcano, Walter says.

Biggs says that the combination of satellite data and AI is a useful tool for drawing attention to possible risks and prioritizing the installation of ground-based instruments. Such remote-monitoring techniques provide valuable information and are safer for scientists, but she thinks they are never going to completely replace having instruments close to the volcano itself.

In the United States, researchers will soon gain a large new source of ground-based data. In March 2019, US legislators passed a bill to fund the National Volcano Early Warning System (NVEWS). When implemented, NVEWS will

lead to the installation of digital broadband seismometers on 104 of the country’s volcanoes and new digital-telemetry networks with sufficient bandwidth to carry data from a number of different ground sensors.

Future shocks

In the past 40 years, scientists have successfully forecast the timing of many eruptions, from smaller blasts at Mount St Helens in the early 1980s to the ash-rich lava fountains at Mount Etna. “A lot of progress has been made on the timing aspect,” Poland says. “Perhaps in a very large part because of the amount of instrumentation, the advent of space-based monitoring and the increase in observations that we have.”

Nevertheless, volcanic eruptions still take people fatally by surprise. A small explosive eruption at Mount Ontake in Japan in 2014 killed 63 people, and a violent eruption of the Fuego volcano in Guatemala in June 2018 killed hundreds. A minor eruption at White Island in New Zealand in 2019 claimed 21 lives.

One challenge facing volcanologists is that they are trying to infer what’s happening deep underground by looking at data such as gas emissions and shape changes on the surface. And each volcano has its own personality — its own unique set of materials and structure.

The individualistic nature of volcanoes highlights the limitations of using patterns from past eruptions to forecast future ones. When volcanologists see the first warning signs, they often think they’ve seen this before and know what happens, Poland says. “But the volcanoes haven’t watched that movie,” he says. “They’ve evolved in ways that are incredibly complex, and our understanding of the complexities are very cursory at this point.”

With more data and better understanding of volcanic systems, researchers hope to develop dynamic models that can capture the physics and chemistry of what happens below ground. In this way, the development of volcanology could parallel that of meteorology, which uses dynamic models of the atmosphere to forecast weather many days in advance.

But volcanic systems are so complex and so hidden that volcanic forecasts will never be as good as meteorological ones, says Poland. “It is a fun exercise to think that one day you will open the newspaper and see the volcano forecast next to the weather forecast,” he says. “But we are still a long way from that.”

Jane Palmer is a freelance writer based in Colorado.

1. Aiuppa, A. et al. *Geophys. Res. Lett.* **37**, L17303 (2010).
2. Ripepe, M. et al. *J. Geophys. Res. Solid Earth* **123**, 9570–9585 (2018).
3. Walter, T. R. et al. *Nature Commun.* **10**, 4339 (2019).
4. Reath, K. et al. *J. Geophys. Res. Solid Earth* **124**, 195–218 (2018).
5. Anantarasirichai, N., Biggs, J., Albino, F. & Hill, P. *J. Geophys. Res. Solid Earth* **123**, 6592–6606 (2018).

John Houghton

(1931–2020)

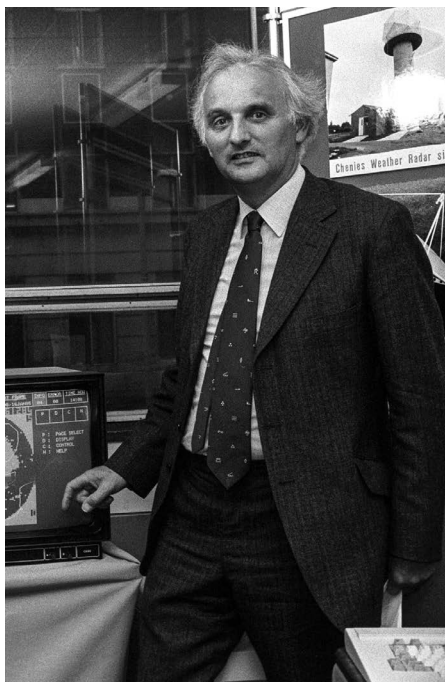
An IPCC founder and tenacious advocate for climate action.

John Houghton was instrumental in founding and shaping the Intergovernmental Panel on Climate Change (IPCC). The climate scientist led the panel's Scientific Assessment of Climate Change working group from its formation in 1988 until 2002. Under his guidance, the IPCC did more than any other entity to synthesize the science, sound the alarm of dangerous climate risk and make the case for immediate action, work for which the organization was awarded the Nobel Peace Prize in 2007. He died on 15 April, aged 88.

In 1990, the IPCC published the first comprehensive global report documenting how greenhouse gases were affecting climate and society. It provided the basis for the United Nations Framework Convention on Climate Change, a treaty negotiated in 1992 at the Earth Summit in Rio de Janeiro that has guided efforts to “prevent dangerous anthropogenic interference with the climate system” ever since. Subsequent assessments led to the Kyoto Protocol of 1997 and the Paris Agreement of 2015, the world's first comprehensive and most ambitious climate treaty.

Born in December 1931 in Dysersyth UK, Houghton's early ambition was, in his words, to “explore the world and the universe in all possible scientific ways”. At 16, he went to study physics at the University of Oxford, UK, where he earned his doctorate in 1955. There, Houghton initially became prominent in the nascent field of remote sensing, designing instruments to measure atmospheric temperature and composition for four of NASA's Nimbus satellites in the 1970s. As deputy director of the Rutherford Appleton Laboratory near Oxford from 1979, he championed a series of instruments to measure sea surface temperature that have been used to track ocean warming for the past three decades.

Houghton headed the UK Met Office from 1983 to 1991. He became increasingly concerned about the impacts of human emissions on Earth's climate, and helped to set up the IPCC as a joint body of the World Meteorological Organization and the United Nations Environment Programme. He also enlisted the support of UK Prime Minister Margaret Thatcher to create the Met Office's Hadley Centre for Climate Science and Services, one of the world's premier climate research centres. As chair of the Royal Commission on Environmental Pollution from 1992 to 1998,



he led a series of major studies on air pollution, including an influential 1994 report on transport and the environment.

Houghton was renowned for his clear communication of science to decision-makers. At the first meeting of the IPCC report's authors, in 1989, he told the atmospheric scientists they needed to come up with a metric for standardizing emissions measurements. When they argued that it was too difficult, Houghton insisted that they must, because they were the best people in the world to do this. The result remains the gold standard for policy: global warming potential (GWP), which compares the impact of different greenhouse gases on climate.

As the IPCC's work continued to advance with the assessments of 1995 and 2001, Houghton worked to keep the agency impartial. It was “entirely a scientific body, not political”, Houghton recalled at a 2015 talk in Cambridge, UK, to mark the 25th anniversary of the first IPCC report and the fifth edition of his textbook *Global Warming: The Complete Briefing*. All too soon, he said, “there were people there who were political who were trying to get rid of what we were trying to do. So they were very tough meetings indeed.”

Houghton proved to be a consummate diplomat. Obtaining agreement from every country in the world is close to impossible;

he achieved it, again and again. An example of his diplomatic skills is when it was time to approve the summary for the 1995 climate assessment. Several oil-producing countries blocked the unanimous agreement required by the report. Houghton suggested adding footnotes to reflect their objections. The countries agreed. However, when they discovered that they were named in the footnotes, their objections disappeared – as did the footnotes.

Houghton dedicated his life to communicating the science of climate change to everyone from religious leaders to foreign dignitaries because of the dangers that lay ahead. When speaking of the overwhelming challenge, he would immediately connect the dots. “What impact will this have on us, and our lives?” he would ask, interspersing charts and maps with images of real people affected by increasing floods and droughts, heat waves and sea level rises. Ramming home our collective responsibility, he would demand of his audience, “Do any of you have an electric car? Nobody at all? I have an electric car!”

His motivation for explaining these dangers came, in large part, from his Christian faith. He was keenly aware that the world's poorest and most vulnerable would suffer disproportionately. He ended the statement on global warming he wrote for the International Society for Science & Religion in 2017 thus: “three qualities ... should guide our stewardship – honesty, holism ... and humility. The alliteration of the three Hs assists in keeping them in mind.”

Houghton influenced countless climate scientists. He would often send his colleagues copies of his books on science and faith, and signed off on his e-mails with “every blessing”. His life was lost to the coronavirus pandemic, a global issue that requires serious and immediate collective action. He spent his life advocating just such a response to climate change.

Katharine Hayhoe is the Endowed Professor in Public Policy and Public Law at Texas Tech University in Lubbock, Texas. Like Houghton, she has been motivated by her faith to study and communicate the risks of climate change.

Donald Wuebbles is the Harry E. Preble Professor of Atmospheric Sciences at the University of Illinois at Urbana–Champaign. He worked with Houghton on four of the IPCC's scientific assessments and special reports. e-mails: Katharine.hayhoe@ttu.edu; wuebbles@illinois.edu

Robert May

(1936–2020)

Pioneering theoretical ecologist and outspoken government adviser.

Robert (Bob) May was the leading mathematical ecologist of his generation and one of the most influential individuals in UK science. A towering intellect, he combined an extraordinarily quick analytical brain with an ability to synthesize information.

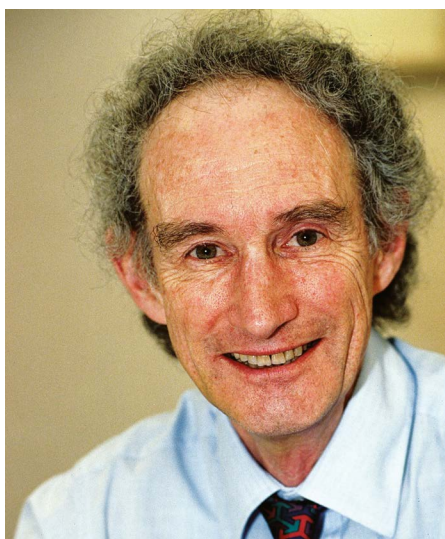
May's contributions were in three main areas: biodiversity, population dynamics and infectious-disease epidemiology. In each, he generated fundamental insights by creating analytical mathematical models that captured the essence of the biology. He helped to transform ecology from a descriptive discipline into a quantitative, analytical science.

From 1995 to 2000, May was chief scientific adviser to the UK government. With his direct style (he never shied away from blunt phrases), May established the role as a high-profile public post, unafraid to speak truth to power. May was forceful on many issues, including bovine spongiform encephalopathy (BSE), genetically modified crops, climate change, homeopathy and infectious diseases. Clarity of thinking and expression enabled him to present complex issues in an understandable way without sacrificing rigour. He developed the UK 'Principles of scientific advice to government', emphasizing three principles: transparency, seeking a wide range of views and fully acknowledging uncertainties. These are particularly relevant in the current pandemic, during which scientists have been called on to advise on policy to deal with a new and poorly understood threat.

May was born in Sydney, Australia, in 1936. He switched from studying chemical engineering at the University of Sydney to physics, attracted to its analytical nature. He completed his PhD in the field of superconductivity. In the 1950s, theoretical physicists were beginning to rely on computers to solve equations; May preferred pencil and paper.

In 1969, motivated by a group at Sydney concerned with social responsibility of scientists and influenced by the renowned ecologist Charles Birch, he became interested in the factors that would prevent collapse in ecosystems such as coral reefs. Two years later, May met Robert McArthur, then the world's leading theoretical ecologist. Following McArthur's untimely death, May moved to Princeton University in New Jersey to succeed him as Class of 1877 Professor of Zoology.

The prevailing view in the 1960s was that ecosystems with little biological diversity,



such as agricultural monocultures, were more unstable because, for example, disease could spread rapidly through them. In his 1973 book *Stability and Complexity in Model Ecosystems*, May showed that communities of competing species become less stable as diversity increases, unless mechanisms that promote stability affect their interactions. The research his models stimulated has since demonstrated such mechanisms, for instance, plants extracting nutrients from the soil at different depths. This is highly relevant in current debates about the resilience of ecosystems to climate change and other disruptions.

May also became one of the pioneers of applying chaos theory to biology. Population ecologists seek to understand why the number of individuals in a population changes over generations. May showed that, for populations with discrete generations, the same model (a non-linear difference equation) could produce stability, cycles of increase and decrease, or chaotic fluctuations, depending on its parameters. The slightest change in initial conditions could lead to widely divergent patterns.

In 1988, May relocated to England as a Royal Society Research Professor, based at the University of Oxford, UK, and Imperial College, London. With the epidemiologist Roy Anderson, May developed a series of insightful analytical models, summarized in their 1991 book *Infectious Diseases of Humans: Dynamics and Control*. Their key innovation was reducing the problem of understanding why and when diseases spread to a few key variables. If, for example, the number of new infections from

one primary case (the transmission factor, R_0) exceeds one, the disease has the potential to become an epidemic. Anderson and May calculated the effective transmission factor if a fraction of the population is immune, for instance as a result of vaccination. This allowed them to predict the proportion of the population that would need to be vaccinated to prevent the spread of a disease. These insights form the foundation of our understanding of the coronavirus pandemic, as R_0 has moved from technical papers into news bulletins around the world.

Between 2000 and 2005, May was president of the Royal Society. There, he particularly championed female scientists, leading to an increase in the number of women elected to the fellowship. And as an independent member of the second chamber of the UK Parliament, he was an effective inquisitor of government ministers as a part of the House of Lords Science and Technology Select Committee.

The 2008 financial crisis led May in a new direction. Building on his work on the stability of ecological communities, May and his colleagues analysed the banking system. The insights generated by his modelling included recommendations, now paying dividends, that the interconnectedness of the banking system should be limited and that the capital reserves of banks should be increased to promote greater stability.

Bob was a loyal colleague, and an extremely competitive individual (it was once said that when he went home to play with his much-loved poodle Perri, he played to win). He was a keen runner – one of us, J.R.K., ran more than 15,000 kilometres with May over a period of 20 years. An annual highlight for Bob was the summer walk for his academic colleagues that he organized for over 40 years, starting in 1975. For a few days, at beautiful mountainous locations around Europe, and accompanied by his wife Judith, Bob would lead a heady mix of camaraderie and scientific discussion.

John R. Krebs is emeritus professor of zoology at the University of Oxford, UK, and was founding head of the UK Food Standards Agency, created following the UK BSE crisis.

Michael Hassell is professor of insect ecology and honorary principal research fellow, department of life sciences at Imperial College London.
e-mail: john.krebs@zoo.ox.ac.uk

Correspondence

Carbon tax to aid economic recovery

The fall in fossil-fuel prices offers governments a chance to offset the potentially massive public debt incurred by the COVID-19 pandemic. Large revenues could be raised by placing a global price on carbon while oil prices are low. For oil alone, a levy of US\$30 per barrel on 100 million barrels per day would return \$3 billion per day, or \$1.1 trillion per year.

A consistent charge applied to all fossil-carbon emissions, irrespective of source, could return overall fossil-fuel prices to pre-pandemic levels in a simple, efficient and transparent way. As well as raising revenue, it would reduce uncertainty about energy prices, buffering against future price spikes and protecting investments in energy efficiency and renewable energy. The economic recovery from the pandemic would follow a low-carbon trajectory.

If world leaders act fast, using the same decisive and coordinated approach they have applied to combat the spread of the virus, they can help to protect both the economy and the climate through a single simple instrument.

Eric Galbraith, Jeroen van den Bergh Institute of Environmental Science and Technology, Autonomous University of Barcelona, Spain.
eric.d.galbraith@gmail.com

Antibody production to bypass animals

The European Commission's Joint Research Centre has just released its recommendations on non-animal-derived antibodies (see go.nature.com/2ypgstg), in accordance with the EU's 2010 directive on protecting laboratory animals (go.nature.com/2wxd9as). We urge government authorities, funding agencies and publishers to endorse this technical advance to improve scientific reproducibility and benefit society.

Animal-derived antibodies are plagued by efficacy issues (A. Bradbury and A. Plückthun *Nature* **518**, 27–29; 2015), with repercussions for research reproducibility, diagnosis and health management. By contrast, non-animal antibodies derived from universal naive display libraries (see, for example, P. Mondon *et al. Front. Biosci.* **13**, 1117–1129; 2008) are defined by sequence and so are consistently reproducible.

Such libraries contain an enormous repertoire of structurally diverse antibody genes, comparable to those of the naive immune system. This facilitates the selection of antibodies for specificity, stability, yield and affinity. The libraries can also be used repeatedly, unlike recombinant animal-derived ones, which require a new immunization protocol for each antigen under investigation. Non-animal antibodies can be engineered in immunoglobulin formats to have properties that are indistinguishable from those of animal-derived ones. They are therefore able to replace them in all known applications.

Alison C. Gray* University of Nottingham, UK.
draligray@live.com

*On behalf of 6 correspondents; see go.nature.com/2atk3cd

Boost longevity of economic model

The COVID-19 pandemic has led economists to weigh up the 'dollar value' of human lives against the effect of lockdown measures on gross domestic product – in order to justify lifting them. Given the burgeoning medical crisis, we should instead be questioning the wisdom of persisting with an economic model that cannot survive a pause of even a few months.

The pandemic presents an opportunity to rethink the rationality of our socio-economic model and to replace it with a more resilient one. A system that relies on complex webs of growing debt, and that ultimately endorses the ever-increasing use of finite physical resources, is by definition unsustainable, even without pandemics. We also need to build in the near-certain emergence of other lethal pathogens in the future (see, for example, K. E. Jones *et al. Nature* **451**, 990–993; 2008).

Scientists and scientific organizations have a responsibility to clearly communicate these long-term considerations to policymakers if such goals are to be realized.

Georgi K. Marinov Stanford University School of Medicine, Stanford, California, USA.
marinovg@stanford.edu

COVID-19: lessons from Lombardy

Since becoming Europe's first epicentre for COVID-19, Lombardy in northern Italy has been a testing bench for managing the coronavirus. Last month, its Regional Forum for Research and Innovation (go.nature.com/3fmtupu) issued recommendations for responsible governance in those areas. Others might also find these recommendations useful during the pandemic.

I write on behalf of the forum, which is an independent advisory board. As well as emphasizing the importance of citizens' participation in creating practical solutions to the crisis and its aftermath, it strongly recommends clarifying to the general public the role and limits of the public-health data on which policymakers' decisions are based – just as data processed through artificial intelligence and algorithms must be openly scrutinized.

The forum also encourages contributions from voluntary human resources, citizen-science initiatives and social innovation organizations – for example, information on and support for the socio-economic and psychological impacts of lockdown measures against the virus. To help speed up resolution of the public-health crisis, the forum advises governments – and not just those in Lombardy – to coordinate and sustain the actions of organizations that can provide such collaborations.

Angela Simone Giannino Bassetti Foundation, Milan, Italy.
angela.simone@fondazionebassetti.org

News & views

Medical research

Statin drugs might boost healthy gut microbes

Peter Libby

An analysis of faecal samples reveals that obese people who take cholesterol-lowering statin drugs have a 'healthier' community of gut microorganisms than would be expected. What are the implications of this surprising finding? **See p.310**

Our digestive systems harbour more bacterial cells than there are human cells in our bodies. Although the often-mentioned estimate of a tenfold excess of microorganisms over human cells might exaggerate the ratio¹, even conservative estimates² accord the microbes numerical dominance at a ratio of about 1.3:1. These close gut microbial neighbours of ours comprise around 0.3% of a person's mass², and there are more than 100 times more bacterial genes in the gut³ than there are genes in their human host. Interest has burgeoned in the potential effects of these normal gut residents (sometimes termed commensal bacteria) on our well-being. On page 310, Vieira-Silva *et al.*⁴ report an unexpected discovery, regarding patterns of gut microbes, that might have clinical consequences.

When trying to assess the daunting complexity of the many thousands of bacterial species in our gut, one option available is a categorization method⁵ that assigns an individual's microbial profile to one of four groupings called enterotypes (Fig. 1), depending on the abundance of signature species. Those of us not immersed in the world of bacterial binomials owe a debt of gratitude to the colonically oriented colleagues who came up with this and other possible classification approaches.

The 'dysbiotic' enterotype called *Bacteroides* 2 (Bact2) is associated with inflammation⁶. People who have this enterotype tend to have a lower load of gut microbes than do those with other enterotypes, and more *Bacteroides* bacteria than *Faecalibacterium* microbes. These individuals also have a higher blood concentration of C-reactive protein – a hallmark of inflammation – than do individuals who have other enterotypes⁷.

A cascade of data from the colonic

cognoscenti links the composition of gut microbes to aspects of health. For example, more than 75% of individuals who have inflammatory bowel disease have the Bact2 enterotype, whereas fewer than 15% of people who do not have the disease harbour this enterotype⁶. Beyond the gut, many researchers have implicated gut microbes in obesity⁸ and the cluster of conditions referred to as metabolic syndrome. However, the nature of the relationship between microbes and these conditions remains under debate.

Studies have also linked gut bacteria to cardiovascular disease. Molecules such as

trimethylamine oxide, which are made by gut bacteria, might accelerate atherosclerosis, and their presence is associated with adverse cardiovascular outcomes, including death⁹. Vieira-Silva *et al.* report that the Bact2 mix of intestinal bacteria is characterized by a paucity of bacterial producers of another microbial molecule, butyrate. This short-chain fatty acid might help to preserve the barrier function of the epithelial cells that line the gut, perhaps preventing leakage of harmful bacterial endotoxin molecules from the bowel and thereby dampening systemic inflammation of the body¹⁰.

In their quest for a potential connection between the bacterial population of the gut and obesity, Vieira-Silva and colleagues made a striking discovery when they mined data collected in a European Union study called the MetaCardis project (<http://www.metacardis.net>). This project has gathered data on the composition of human gut microbes using state-of-the-art technology, to assess the microbes' role in cardiovascular disease. More than 2,000 individuals recruited from European countries took part in an exhaustive survey that collected data for around 1,400 variables, such as medication taken and body-mass index (a measure used to assess a person's weight that takes height into account).

Vieira-Silva *et al.* report that in a subset

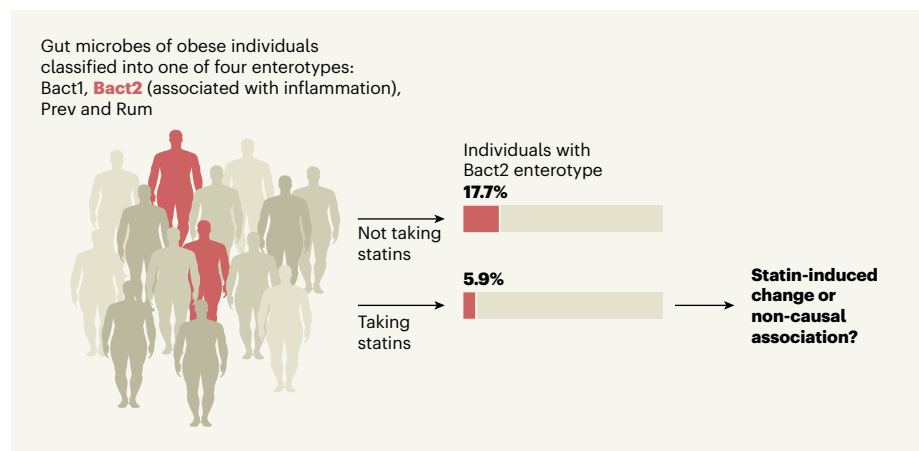


Figure 1 | Gut-microbe changes associated with the use of statin drugs. A person's gut microorganisms can be classified⁵, by the analysis of faecal samples, into one of four groups called enterotypes, depending on the abundance of particular microbial species. These groupings are termed *Bacteroides* 1 (Bact1), *Bacteroides* 2 (Bact2), *Ruminococcaceae* (Rum) and *Prevotella* (Prev). The Bact2 enterotype is associated with health problems and inflammation⁶. Vieira-Silva *et al.*⁴ assessed enterotype data for individuals who were recruited as part of a project to understand factors influencing cardiovascular health. The authors made the unexpected discovery that the prevalence of the Bact2 enterotype was lower than expected in obese individuals who were taking cholesterol-lowering drugs called statins. Whether this decrease in prevalence of the Bact2 enterotype in obese individuals is directly caused by statins or is due to another factor associated with statin use (if individuals taking statins have better access to health care, for example) will require further study.

of nearly 900 participants whose data they analysed, a higher prevalence of the Bact2 enterotype correlated with a higher body-mass index and obesity. However, the authors made the striking discovery that the pattern of enterotypes found in the population of obese individuals differed significantly depending on whether people were taking cholesterol-lowering drugs called statins (comprising about 12% of those studied). This result raised a surprising possible connection between statin intake and gut microbes. The obese participants taking statins had a significantly lower prevalence of the Bact2 enterotype (5.9% of the obese population) than did their obese counterparts not taking statins (17.7% of the obese population). Vieira-Silva and colleagues confirmed this phenomenon in an independent data set from the Flemish Gut Flora Project¹¹.

The use of statins is one of the great success stories of modern cardiovascular therapeutics. Originally derived from natural products of microbial denizens of the soil, these agents inhibit a rate-limiting enzyme in the pathway that makes cholesterol. By lowering cholesterol production, the treatment coaxes cells to boost the expression of receptors for low-density lipoprotein (LDL) that capture cholesterol-rich LDL particles, and this results in a robust decrease in cholesterol in the bloodstream. This LDL reduction substantially lowers the risk of cardiovascular events such as heart attack and stroke in a large swathe of the population at risk of such conditions, and many people use drugs of the statin class. Large meta-analyses of the effects of statin treatment reveal that it prolongs lifespan and that, on balance, the benefits outweigh any unwanted effects¹².

Independently of their effects on LDL, statins have anti-inflammatory actions that probably contribute to their clinical benefit through well-established molecular mechanisms¹³. However, no statin study has singled out obese individuals as targets for therapy, and no current guideline recommends considering obesity when making decisions about using statins for treatment.

Vieira-Silva and co-workers' unexpected findings therefore raise intriguing questions relating to the clinical use of statins. Yet interpretation of these findings warrants caution, in particular with regard to the risk of confusing correlation with causation. As the authors of this large and carefully executed study rightfully acknowledge, we should consider whether statin takers have had better access to health care or been more engaged in other health-promoting behaviours than have the individuals who were not taking statins. A large-scale clinical trial to determine whether statins lead to a reduced prevalence of the Bact2 enterotype in obese participants who would not otherwise receive statins could

address this possibility, which is known as confounding by indication. Moreover, whether these findings apply across ethnic groups will require further study. In any case, following up on these provocative observations promises to provide new mechanistic insight into the complex relationships between obesity, metabolic status, gut microbes and cardiovascular disease.

Peter Libby is in the Department of Medicine, Division of Cardiovascular Medicine, Brigham and Women's Hospital, Harvard Medical School, Boston, Massachusetts 02115, USA. e-mail: plibby@bwh.harvard.edu

Applied physics

Artificial eye boosted by hemispherical retina

Hongrui Jiang

An artificial eye has been reported that incorporates densely packed, nanometre-scale light sensors into a hemispherical retina-like component. Some of its sensory capabilities are comparable to that of its biological counterpart. **See p.278**

Science fiction frequently features robots that have artificial eyes, as well as bionic eyes that interface with the human brain to restore the vision of people who are blind. Much effort has been made to develop such devices, but fabricating the spherical shape of a human eye – particularly a hemispherical retina – is an enormous challenge that severely limits the function of artificial and bionic eyes. On page 278, Gu *et al.*¹ report an innovative, concavely hemispherical retina consisting of an array of nanometre-scale light sensors (photosensors) that mimic the photoreceptor cells in human retinas. The authors use this retina in an electrochemical eye that has several capabilities comparable to those of the human eye, and that performs the basic function of acquiring image patterns.

The human eye, with its hemispherical retina, has a more ingenious optical layout than, say, that of the flat image sensors in cameras: the dome shape of the retina naturally reduces spreading of light that has passed through the lens, thus sharpening the focus. The core component of Gu and colleagues' biomimetic electrochemical eye is the high-density array of photosensors that serves as the retina (Fig. 1). The photosensors were formed directly inside the pores of a hemispherical membrane of aluminium oxide (Al₂O₃).

Thin, flexible wires made of a liquid metal

1. Nature <https://doi.org/10.1038/nature.2016.19136> (2016).
2. Sender, R., Fuchs, S. & Milo, R. *PLoS Biol.* **14**, e1002533 (2016).
3. Qin, J. *et al.* *Nature* **464**, 59–65 (2010).
4. Vieira-Silva, S. *et al.* *Nature* **581**, 310–315 (2020).
5. Arumugam, M. *et al.* *Nature* **473**, 174–180 (2011).
6. Vieira-Silva, S. *et al.* *Nature Microbiol.* **4**, 1826–1831 (2019).
7. Costea, P. I. *et al.* *Nature Microbiol.* **3**, 8–16 (2018).
8. Graham, C., Mullen, A. & Whelan, K. *Nutr. Rev.* **73**, 376–385 (2015).
9. Tang, W. H. W., Bäckhed, F., Landmesser, U. & Hazen, S. L. *J. Am. Coll. Cardiol.* **73**, 2089–2105 (2019).
10. Baxter, N. T. *et al.* *mBio* **10**, e02566-18 (2018).
11. Falony, G. *et al.* *Science* **352**, 560–564 (2016).
12. Collins, R. *et al.* *Lancet* **388**, 2532–2561 (2016).
13. Oesterle, A., Laufs, U. & Liao, J. K. *Circ. Res.* **120**, 229–243 (2017).

The author declares competing financial interests: see go.nature.com/3anodse for details.

This article was published online on 6 May 2020.

(eutectic gallium–indium alloy) sealed in soft rubber tubes transmit signals from the nanowire photosensors to external circuitry for signal processing. These wires mimic the nerve fibres that connect the human eye to the brain. A layer of indium between the liquid-metal wires and nanowires improves electrical contact between the two. The artificial retina is held in place by a socket made from a silicone polymer, to ensure proper alignment between the wires and nanowires.

A lens combined with an artificial iris is placed at the front of the device, just as in the human eye. The retina at the back combines with a hemispherical shell at the front to form a spherical chamber (the 'eyeball'); the frontal hemispherical shell is made from aluminium lined with a tungsten film. The chamber is filled with an ionic liquid that mimics the vitreous humour – the gel that fills the space between the lens and the retina in the human eye. This arrangement is necessary for the electrochemical operation of the nanowires. The overall structural similarity between the artificial eye and the human eye confers on Gu and colleagues' device a wide field of view of 100°. This compares with roughly 130° for the vertical field of view of a static human eye.

The structural mimicry of Gu and colleagues' artificial eye is certainly impressive, but what makes it truly stand out from previously reported devices is that many of its

sensory capabilities compare favourably with those of its natural counterpart. For example, the artificial retina can detect a large range of light intensities, from 0.3 microwatts to 50 milliwatts per square centimetre. At the lowest intensity measured, each nanowire in the artificial retina detects an average of 86 photons per second, on a par with the sensitivity of photoreceptors in human retinas. This sensitivity derives from the perovskite material used to make the nanowires. Perovskite compounds are extremely promising materials for various optoelectronic and photonic applications². The perovskite used by Gu *et al.* is formamidinium lead iodide, and was chosen for its excellent optoelectronic properties and good stability.

The responsivity of the nanowires, which measures the current produced per watt of incident light, is almost the same for all frequencies of the visible spectrum. Moreover, when the nanowire array is stimulated by regular, rapid pulses of light, it can produce a current in response to a pulse in just 19.2 milliseconds, and can then take as little as 23.9 ms to recover (return to its inactive state) when the pulse has ended. The response and recovery times are important parameters, because they ultimately determine how quickly the artificial eye can respond to a light signal. For comparison, the response and recovery times of photoreceptors in human retinas range from 40 to 150 ms.

Perhaps most impressive is the high resolution of the imaging achieved by Gu and colleagues' artificial retina, which results from the high density of the nanowire array. In previous artificial retinas, the photosensors were first fabricated on flat, rigid substrates; after that, either they were transferred onto curved supporting surfaces³ or the substrate was folded into a curve⁴. This limited the density of the imager units, because space had to be left between them to allow for the transfer or folding.

By contrast, the nanowires in Gu and co-workers' device are formed directly on a curved surface, which allows them to be packed together more closely. Indeed, the nanowire density is as high as $4.6 \times 10^8 \text{ cm}^{-2}$, much greater than that of photoreceptors in the human retina (about 10^7 cm^{-2}). The signal from each nanowire can be acquired individually, but the pixels in the current device were formed from groups of three or four nanowires.

The overall performance of Gu and colleagues' artificial eye represents a leap forwards for such devices, but much still needs to be done. First, the photosensor array is currently only 10×10 pixels, with roughly 200- μm gaps between the pixels; this means that the light-detecting region is only about 2 mm wide. Moreover, the fabrication process involves some costly and low-throughput

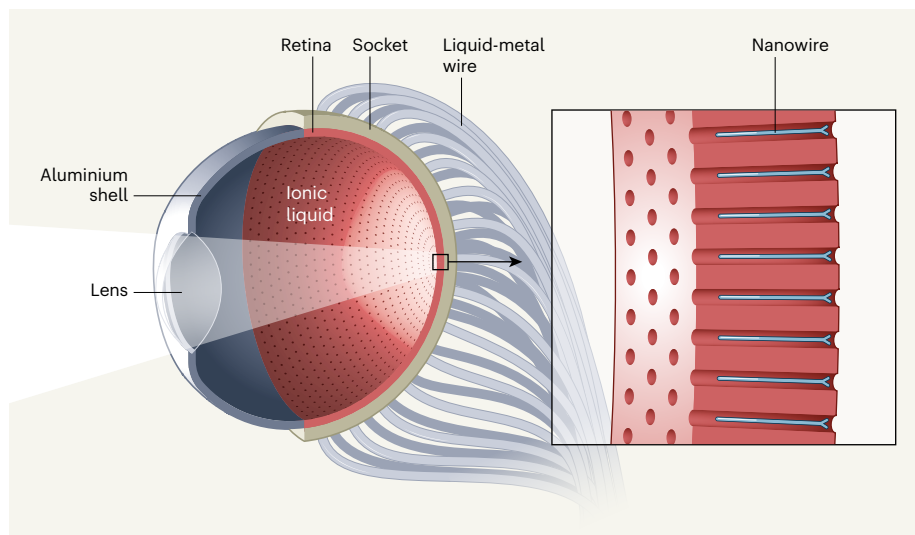


Figure 1 | A biomimetic artificial eye. Gu *et al.*¹ report an artificial visual system that mimics the human eye. A lens is fixed over an aperture in an 'eyeball', which consists of a metal shell at the front, an artificial retina at the back and an ionic liquid in the middle. The key advance is the hemispherical retina: a dense array of light-sensitive nanowires held in the pores of an aluminium oxide membrane. The nanowires mimic the photoreceptor cells in biological retinas. A polymeric socket holds the retina, ensuring electrical contact between the nanowires and liquid-metal wires at the back. The liquid-metal wires mimic the nerve fibres by transmitting signals from the nanowires to external circuitry for signal processing.

steps – for example, an expensive process known as focused-ion-beam etching is used to prepare each pore for nanowire formation. High-throughput fabrication methods must be developed in the future to produce larger photosensor arrays, at drastically reduced cost.

Second, to improve the resolution and scale of the retina, the size of the liquid-metal wires will need to be reduced. The outer

“Perhaps most impressive is the high resolution of the imaging achieved by the artificial retina.”

diameter of the wires is about 700 μm , but this should ideally be comparable to the nanowire diameter (a few micrometres). It is currently challenging to reduce the diameter of the liquid-metal wires to that size.

Third, more testing is needed to establish the operational lifetime of the artificial retina. Gu *et al.* report that there is no obvious reduction in its performance after nine hours of operation, but the performance of other electrochemical devices can deteriorate over time. Lastly, the authors note that the response and recovery times of their device are reduced at higher concentrations of the ionic liquid, but at the expense of light transmission through the liquid. Further optimization of the ionic-liquid composition is needed to address this problem.

Nevertheless, Gu and colleagues' work adds to the breakthroughs that have been made

in the past few decades^{3–9}, which have been achieved by mimicking not only camera-like eyes (such as those of humans), but also compound eyes similar to those of insects. Given these advances, it seems feasible that we might witness the wide use of artificial and bionic eyes in daily life within the next decade.

Hongrui Jiang is in the Department of Electrical and Computer Engineering, the Department of Materials Science and Engineering, the Department of Biomedical Engineering, the Department of Ophthalmology and Visual Sciences, and the McPherson Eye Research Institute, University of Wisconsin – Madison, Madison, Wisconsin 53706, USA.
e-mail: hongrui@engr.wisc.edu

1. Gu, L. *et al.* *Nature* **581**, 278–282 (2020).
2. Stranks, S. D., Snaith, H. J. *Nature Nanotechnol.* **10**, 391–402 (2015).
3. Ko, H. C. *et al.* *Nature* **454**, 748–753 (2008).
4. Zhang, K. *et al.* *Nature Commun.* **8**, 1782 (2017).
5. Jung, I. *et al.* *Proc. Natl Acad. Sci. USA* **108**, 1788–1793 (2011).
6. Liu, H., Huang, Y., Jiang, H. *Proc. Natl Acad. Sci. USA* **113**, 3982–3985 (2016).
7. Floreano, D. *et al.* *Proc. Natl Acad. Sci. USA* **110**, 9267–9272 (2013).
8. Huang, C. C. *et al.* *Small* **10**, 3050–3057 (2014).
9. Jeong, K. H., Kim, J., Lee, L. P. *Science* **312**, 557–561 (2006).

Fourth defence molecule completes antiviral line-up

Niklas A. Schmacke & Veit Hornung

Toll-like receptors can initiate an immune response when they detect signs of a viral or microbial threat. New insight into how such receptor activation drives defence programs should aid our efforts to understand autoimmune diseases. **See p.316**

Cells of the innate branch of the immune system can detect infectious agents outside the cell or in various intracellular compartments. This process depends on the sensing of 'non-self' molecular signatures by proteins known as pattern-recognition receptors (PRRs)¹. In the endolysosome, an organelle into which extracellular material can be taken up by the cell, PRRs called Toll-like receptor 7 (TLR7), TLR8 and TLR9 can be activated by the presence of viral or microbial nucleic acids². However, these same receptors are often linked to the erroneous detection of 'self' nucleic acids in autoimmune diseases³. Heinz *et al.*⁴ report on page 316 that a protein they name TASL links the activation of TLR7, TLR8 and TLR9 to the production of molecules called type I interferons, which mediate antiviral defence. The gene that encodes TASL has been associated with the autoimmune disease systemic lupus erythematosus, and this finding might shed light on factors that contribute to the disease.

To distinguish between different disease-causing viruses and microbes and to tailor a suitable response, the innate immune system uses PRRs in various parts of the cell¹. Each of these sensors recognizes a distinct hallmark of infectious agents termed a pathogen-associated molecular pattern (PAMP). One such family of receptors, the TLRs (Fig. 1), are transmembrane proteins that are found on the cell surface or in the endolysosome². Most cell-surface TLRs detect bacterial components, such as lipopeptides found in bacterial cell walls. By contrast, the endolysosomal TLRs – TLR3, TLR7, TLR8 and TLR9 – recognize nucleic acids or their degradation products, which are typically associated with viral infection but are also a signature of living microbes.

On activation in response to binding a PAMP, TLRs engage another protein, termed an adaptor protein, which provides a crucial control point that sets off distinct signalling cascades culminating in defence responses². Together with other defence mechanisms,

two major gene-expression programs can be distinguished that are broadly tailored to the particular threat sensed. Downstream of most TLRs, an adaptor protein called MyD88 activates the transcription-factor protein NF- κ B, which drives expression of pro-inflammatory genes as part of the immune response. A subgroup of TLRs (TLR3 and TLR4) can engage the protein TRIF, which acts as a scaffold enabling a kinase enzyme to add a phosphate group to the transcription factor IRF3. This phosphorylation activates IRF3, a member of a family of transcription factors

termed interferon regulatory factors (IRFs), which activate broad gene-expression programs. A hallmark of these programs is the production of type I interferon molecules⁵.

Interferons are potent drivers of a branch of the immune system termed the adaptive immune response, and their presence therefore runs the risk of contributing to autoimmunity. To prevent such an attack by the host's own immune system, an interferon response must be tightly regulated. As a safeguard, a particular sequence of amino-acid residues in TRIF, the pLxIS motif, must be phosphorylated before IRF3 can be activated. This control mechanism provides a 'licensing step' that is not specific just for TRIF as an adaptor protein for TLR signalling, but is a general hallmark of sensing pathways that engage IRF3, or the related protein IRF7, to drive interferon expression. Every identified innate sensing pathway connecting the recognition of nucleic acids to the production of type I interferons, with one exception, had been shown previously to signal through one of the three adaptor proteins known so far to contain a pLxIS motif: TRIF, MAVS and STING. Thus, pLxIS-motif-containing adaptor proteins specifically hardwire nucleic-acid recognition to antiviral defences.

The only exception to this rule had been the endolysosomal TLRs – TLR7, TLR8 and

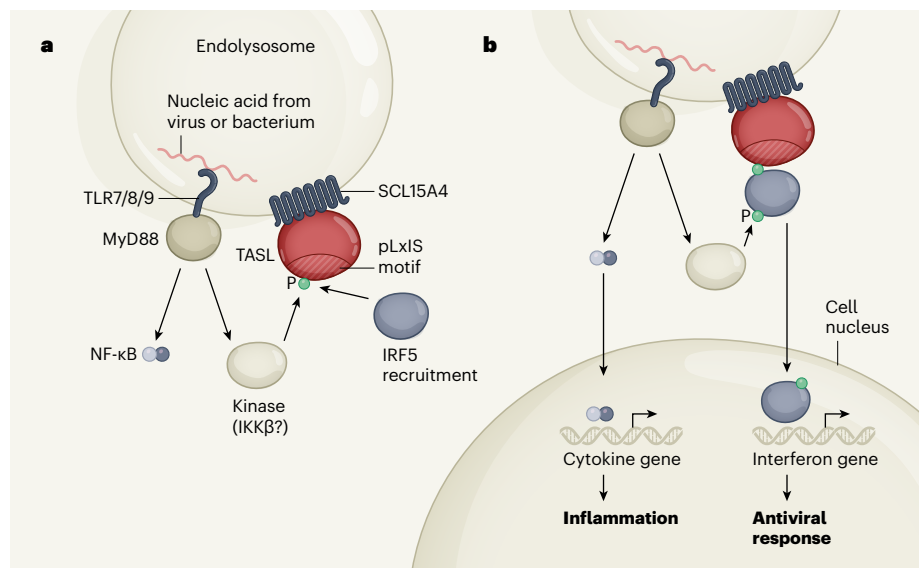


Figure 1 | A crucial role for the TASL protein in the activation of interferon-mediated antiviral defences.

a, Nucleic acids from viruses or bacteria can be taken up by human immune cells into an organelle called the endolysosome and are recognized by Toll-like receptor proteins (TLR7, TLR8 and TLR9). This recognition activates the protein MyD88, which, in turn, activates the transcription-factor protein NF- κ B, a key player in the immune response to infection. Heinz *et al.*⁴ investigated how a protein called SCL15A4, which is associated with the autoimmune disease systemic lupus erythematosus, aids this defence response. They found that it binds to a protein named TASL, which contains an evolutionarily conserved amino-acid sequence called the pLxIS motif. When endolysosomal TLRs recognize foreign nucleic acids, a phosphate group (P) is added to TASL by a kinase enzyme (possibly IKK β from a pathway downstream of MyD88). This phosphorylation recruits the transcription factor IRF5 to TASL. **b**, TASL then acts as a scaffold to facilitate the phosphorylation and activation of IRF5 by a kinase (possibly IKK β). It is the first pLxIS-containing protein known to mediate IRF5 activation. Phosphorylated IRF5 enters the nucleus and drives the expression of genes that encode antiviral interferon molecules. NF- κ B drives the expression of pro-inflammatory defence molecules called cytokines.

TLR9. Although these detect nucleic acids and drive type I interferon gene expression, they do not use TRIF (ref. 2). Instead, they use MyD88 to activate the interferon regulatory factor IRF5 (ref. 6), which is related in structure and function to IRF3 and IRF7. A licensing step involving a pLxIS motif had not been found previously in the signalling cascades of TLR7, TLR8 and TLR9.

Immune cells called plasmacytoid dendritic cells express high levels of TLR7 and TLR9, and are crucial to antiviral defences through their production of large amounts of type I interferon. But these cells are also central players in systemic lupus erythematosus⁷. Previous studies of the molecular mechanisms underlying this disease^{7,8} have identified a protein called SLC15A4, located on endolysosomal membranes, which transports polypeptides and the amino acid histidine. SLC15A4 has been linked to the activation of TLRs.

To investigate the role of SLC15A4 further, Heinz and colleagues used mass spectrometry to probe for proteins that interact with it. This approach identified the protein TASL, which had previously been little researched. TASL is highly abundant in cells of the innate immune system, and Heinz and colleagues found that it is tethered to endolysosomes through interactions with SLC15A4. The authors' further experiments confirmed that this interaction is specific: SLC15A4 bound TASL in immunoprecipitation tests; however, neither the related protein SLC15A3 nor a mutant version of SLC15A4 interacted with TASL in such assays.

When the authors engineered plasmacytoid dendritic cells and immune cells called monocytes to lack expression of the gene encoding TASL, they found that signalling mediated by TLR7, TLR8 and TLR9 was abolished, and a similar effect was seen when SLC15A4 was absent. Heinz *et al.* went on to demonstrate that TASL acts specifically through IRF5 by finding that the response to TLR7 and TLR9 activation remained intact in immune cells lacking IRF3 or IRF7, but was blocked in cells deficient in TASL or IRF5. However, NF- κ B-mediated signalling was unaffected when the pathway acting through TASL was disrupted. Intriguingly, the authors identified a pLxIS motif in TASL, and found evidence that phosphorylation of this motif – by kinases downstream of MyD88 that are associated with NF- κ B activation – mediates IRF5 activation.

This discovery elevates TASL to membership of an exclusive circle of IRF-activating adaptor proteins containing pLxIS motifs, of which the other members are TRIF, MAVS and STING (ref. 5). These four proteins together control the type I interferon response induced by nucleic-acid sensing, a picture that has now been completed with the discovery of TASL as the missing pLxIS adaptor of TLR7, TLR8 and TLR9 signalling.

Given that TASL signals to IRF5, but not to IRF3 or IRF7, it will be interesting to determine the structural features required for the differential recruitment of IRF-family members to pLxIS-motif-containing proteins. Although the authors performed preliminary experiments to investigate phosphorylation events in this system, phosphorylation of the pLxIS motif in TASL should be investigated in detail to identify the kinase(s) responsible.

Moreover, it will be interesting to sort out how this newly identified signalling pathway operates in relation to activation of the pathway involving MyD88, which is the key adaptor of TLR7, TLR8 and TLR9 signalling⁹. As has been shown for other signalling cascades triggered by TLRs and involving multiple adaptors, it is possible that MyD88-mediated signalling and the pLxIS licensing step involving TASL emanate sequentially from distinct endolysosomal vesicles at different stages of maturation². Although TASL is not involved in NF- κ B activation, the authors found that the expression of certain pro-inflammatory genes was still blocked in TASL-deficient cells, probably because of the associated defect in IRF5 activation. Nonetheless, by offering a way to

dampen interferon-mediated autoimmunity in a way that doesn't block the ability to launch an inflammatory defence response, TASL might prove to be a drug target for treating autoimmune diseases that are fuelled by the engagement of TLR7, TLR8 and TLR9.

Niklas A. Schmacke and **Veit Hornung** are in the Department of Biochemistry and at the Gene Center, Ludwig-Maximilians-Universität München, Munich 81377, Germany.
e-mail: hornung@genzentrum.lmu.de

1. Iwasaki, A. & Medzhitov, R. *Nature Immunol.* **16**, 343–353 (2015).
2. Fitzgerald, K. A. & Kagan, J. C. *Cell* **180**, 1044–1066 (2020).
3. Roers, A., Hiller, B. & Hornung, V. *Immunity* **44**, 739–754 (2016).
4. Heinz, L. X. *et al.* *Nature* **581**, 316–322 (2020).
5. Liu, S. *et al.* *Science* **347**, aad2630 (2015).
6. Schoenemeyer, A. *et al.* *J. Biol. Chem.* **280**, 17005–17012 (2005).
7. Sisirak, V. *et al.* *J. Exp. Med.* **211**, 1969–1976 (2014).
8. Blasius, A. L. *et al.* *Proc. Natl Acad. Sci. USA* **107**, 19973–19978 (2010).
9. Hemmi, H. *et al.* *Nature Immunol.* **3**, 196–200 (2002).

This article was published online on 13 May 2020.

Astronomy

An early start for galactic disks

Alfred Tiley

A powerful radio telescope has peered back through time to observe a galaxy that contained a cold, rotating disk of gas not long after the Big Bang – fuelling the debate about when and how disks first formed in galaxies. **See p.269**

Galaxies are immense, gravitationally bound systems composed of stars, dust, gas and invisible 'dark matter'. Understanding how galaxies have formed and grown over time is essential for a more general view of how matter assembles into large structures – a key piece of the puzzle in our efforts to comprehend the Universe. A crucial step towards this goal is to obtain a clear picture of when disk structures first appeared in galaxies. On page 269, Neeleman *et al.*¹ present observations that reveal a massive, rotating disk of cold gas inside a star-forming galaxy only 1.5 billion years after the Big Bang. This is considerably earlier in cosmic history than the times when previously detected gas disks were found to have existed².

According to our current understanding of cosmology, the earliest large-scale structures in the Universe were spherical

dark-matter 'haloes' that collapsed under their own gravity³. Surrounding gas fell into these haloes, subsequently forming stars and, ultimately, galaxies⁴. Haloes and galaxies are thought to have continued to grow together by hierarchical assembly (merging), and through the further accretion of gas and its conversion to stars⁵. Hierarchical assembly is simple, and is thought to be well understood. However, there is still much debate surrounding the exact pathways by which gas accretion and its assembly into stars occurs, and how it relates to the formation of physical and dynamical structures in galaxies over time.

A key component of this mystery is why some galaxies, such as our own star-forming Milky Way, have physical structures dominated by disks of stars and gas (Fig. 1), whereas other, generally older and more quiescent galaxies do not. The answer is probably



Figure 1 | The dusty spiral galaxy NGC 4414. Many star-forming galaxies contain disks of dust and gas – here, the dust is visible as dark patches and streaks silhouetted against the starlight. Neeleman *et al.*¹ report the observation of another galaxy disk that existed just 1.5 billion years after the Big Bang, considerably earlier than previously reported disks.

intimately linked to each galaxy's history of assembly – specifically, to the relative importance of hierarchical merging (which can either promote or destroy disk growth, depending on the circumstances^{6,7}) and of growth through gas accretion (among other processes).

Gas accretion is thought to occur through either a hot or cold mode. As the names suggest, the main difference in these modes is whether the gas is hot or cold as it falls towards the centre of a dark-matter halo onto a galaxy. The hot mode of accretion results in galaxy disks forming late, because a considerable amount of time is needed for the accreted gas to cool and eventually settle into a disk. In the cold mode of accretion, the gas instead remains cool as it falls into the halo centre, thus allowing more-rapid disk formation⁸.

Determining when disks first emerged in galaxies, and how frequently, should thus provide important insights into how the early assembly of galaxies took place. To do this, disks must be found in progressively more-distant galaxies, so that researchers can probe ever further back in time towards the Big Bang. (The light from more-distant galaxies takes longer to arrive at our Earth-bound telescopes and detectors than does light from closer galaxies, and therefore provides

information about the Universe from further back in time.) This requires extremely sensitive instruments that produce high-resolution data. Modern advances in detector and telescope technology, and in instrument design, have enabled the detection of gas disks in massive galaxies that existed around 3 billion years after the Big Bang².

To extend observations of gas in galaxies to even earlier periods of cosmic history, Neeleman *et al.* used the Atacama Large Millimeter/submillimeter Array (ALMA), one of the most powerful radio telescopes in the world, situated in the Atacama Desert in northern Chile. The researchers detected light emitted from cold gas in a galaxy from around 12.5 billion years ago. By resolving the light to a scale of 1.3 kiloparsecs (about one-sixth of the distance from our Sun to the centre of the Milky Way⁹), they were able to examine the structure and kinematics of the emitting gas in impressive detail. They then used simple but robust analytical models to show that their observations are consistent with the presence of a rapidly rotating gas disk, spatially coincident with the galaxy's stars and dust.

Neeleman and colleagues' results constitute some of the first observational evidence for the existence of cold gas disks in massive galaxies very soon after the Big Bang, directly

establishing that massive gas disks could form 1.5 billion years earlier than previous observations had indicated². The authors' work considerably shifts the observational frontier for the detailed study of spatially resolved gas properties in galaxies to when the Universe was only about one-tenth of its current age.

Their discovery is intriguing when viewed alongside the results of some numerical simulations of galaxy formation, which suggest that disks did not begin to dominate in galaxies of similar mass until the Universe was between 4 billion and 6 billion years old^{10,11}. However, it is consistent with the theoretical expectation that cold-mode accretion should be dominant early in the Universe's history⁸. It also ties in with recent, higher-resolution simulations that have seen disks emerge at earlier cosmic epochs¹².

One limitation of the work, when it comes to constraining our theoretical understanding of galaxy formation or testing the differing predictions of numerical simulations, is that the authors consider only one galaxy. Similar observations of many more galaxies from the same epoch are needed before we can determine whether the galaxy studied is representative of the whole population at that time, or whether it is an outlier. Moreover, although the authors' results seem to speak against hot-mode accretion scenarios for early galaxy growth, their data do not explicitly rule out other ways, besides cold-mode accretion, in which cool gas could be efficiently transported to the centres of haloes – for example, through the merging of galaxies and their haloes⁷. Further observational data are required to resolve this issue. Nevertheless, Neeleman and colleagues' findings will excite astronomers, and open up a new epoch of the Universe's history for the study of early galaxy formation.

Alfred Tiley is at the International Centre for Radio Astronomy Research, University of Western Australia, Perth, Western Australia 6009, Australia.
e-mail: alfred.tiley@uwa.edu.au

1. Neeleman, M., Prochaska, J. X., Kanekar, N. & Rafelski, M. *Nature* **581**, 269–272 (2020).
2. Wisnioski, E. *et al.* *Astrophys. J.* **799**, 209 (2015).
3. Blumenthal, G. R., Faber, S. M., Primack, J. R. & Rees, M. J. *Nature* **311**, 517–525 (1984).
4. Fall, S. M. & Efstathiou, G. *Mon. Not. R. Astron. Soc.* **193**, 189–206 (1980).
5. L'Huillier, B., Combes, F. & Semelin, B. *Astron. Astrophys.* **544**, 68 (2012).
6. Welker, C. *et al.* *Mon. Not. R. Astron. Soc.* **465**, 1241–1258 (2017).
7. Grand, R. J. J. *et al.* *Mon. Not. R. Astron. Soc.* **467**, 179–207 (2017).
8. Kereš, D., Katz, N., Weinberg, D. H. & Davé, R. *Mon. Not. R. Astron. Soc.* **363**, 2–28 (2005).
9. Eisenhauer, F. *et al.* *Astrophys. J.* **597**, L121–L124 (2003).
10. Zavala, J. *et al.* *Mon. Not. R. Astron. Soc.* **460**, 4466–4482 (2016).
11. Dubois, Y. *et al.* *Mon. Not. R. Astron. Soc.* **463**, 3948–3964 (2016).
12. Pillepich, A. *et al.* *Mon. Not. R. Astron. Soc.* **490**, 3196–3233 (2019).

THE HUBBLE HERITAGE TEAM (STSC/AURA/NASA)

A cold, massive, rotating disk galaxy 1.5 billion years after the Big Bang

<https://doi.org/10.1038/s41586-020-2276-y>

Marcel Neeleman^{1✉}, J. Xavier Prochaska^{2,3}, Nissim Kanekar⁴ & Marc Rafelski^{5,6}

Received: 30 October 2019

Accepted: 20 March 2020

Published online: 20 May 2020

 Check for updates

Massive disk galaxies like the Milky Way are expected to form at late times in traditional models of galaxy formation^{1,2}, but recent numerical simulations suggest that such galaxies could form as early as a billion years after the Big Bang through the accretion of cold material and mergers^{3,4}. Observationally, it has been difficult to identify disk galaxies in emission at high redshift^{5,6} in order to discern between competing models of galaxy formation. Here we report imaging, with a resolution of about 1.3 kiloparsecs, of the 158-micrometre emission line from singly ionized carbon, the far-infrared dust continuum and the near-ultraviolet continuum emission from a galaxy at a redshift of 4.2603, identified by detecting its absorption of quasar light. These observations show that the emission arises from gas inside a cold, dusty, rotating disk with a rotational velocity of about 272 kilometres per second. The detection of emission from carbon monoxide in the galaxy yields a molecular mass that is consistent with the estimate from the ionized carbon emission of about 72 billion solar masses. The existence of such a massive, rotationally supported, cold disk galaxy when the Universe was only 1.5 billion years old favours formation through either cold-mode accretion or mergers, although its large rotational velocity and large content of cold gas remain challenging to reproduce with most numerical simulations^{7,8}.

An open question in galaxy evolution is the epoch at which disk galaxies like our Milky Way formed. In our current cosmology paradigm, known as Λ Cold Dark Matter⁹, galaxies are expected to be built up in a hierarchical manner. Gas and dark matter funnel into dark matter halos, merging and condensing into larger structures, precipitating the formation of stars and the growth of the galaxy. However, the physical processes that dominate galaxy formation are still under debate.

In the traditional picture of galaxy formation, the infalling gas is shock-heated to the virial temperature (about 10^6 K for a galaxy with a mass of 10^{12} solar masses, M_\odot) and accretes spherically; the central region then cools and condenses into a rotationally supported disk^{1,2}. Besides this ‘hot-mode’ accretion scenario, numerical simulations predict an alternative scenario in which gas accretes efficiently onto galaxies either through the merging of galaxies or through gas flowing directly into galaxies along filamentary structures, with a considerable fraction of the gas remaining cool, at temperatures far below the virial temperature of the galaxies^{3,4}. Unlike the hot-mode scenario, in which the long cooling times imply that disk galaxies form at relatively late times (redshift $z \lesssim 1$), in these latter scenarios, disk galaxies can form much earlier ($z \lesssim 5$)^{7,8}. Observing the earliest onset of galaxy disks can therefore inform us how galaxies acquire their mass, allowing us to distinguish between these mass accretion scenarios.

Observationally, disks have been identified via carbon monoxide (CO) or H α spectroscopy at $z \lesssim 2.5$ (refs. ^{10–12}). At higher redshifts, $z \approx 4–5$, observations with the Karl G. Jansky Very Large Array (JVLA) and

Atacama Large Millimeter/submillimetre Array (ALMA) have yielded tentative evidence of disks^{5,6}. However, a combination of relatively low resolution and sensitivity has meant that it has been impossible to conclusively identify rotating disk galaxies at $z \gtrsim 3$. The vast majority of these studies have focused on objects with high star-formation rates (SFR), selected by their high luminosity in either the optical/near-infrared and/or at submillimetre wavelengths. Because the star-formation efficiency is higher in dense, clumpy environments, such ‘emission-selected’ galaxy samples could be biased against the presence of stable disks. A complementary approach to identifying high-redshift galaxies lies in detecting absorption lines from gas in the galaxy, if the galaxy happens to lie in front of a bright background source such as a quasi-stellar object (QSO). This approach does not contain a bias towards more luminous galaxies, and therefore ‘absorption-selected’ galaxy samples provide a unique luminosity-unbiased sample to understand disk galaxy formation at high redshifts.

Our recent ALMA searches for absorption-selected galaxies using the fine-structure line of singly ionized carbon at $157.74 \mu\text{m}$ ([C II]) have revealed a sample of six galaxies at $z \approx 4$ (refs. ^{13,14}), with SFR of about $(7–110) M_\odot \text{ yr}^{-1}$, as determined from their [C II] and dust continuum emission. Owing to the coarse angular resolution of these observations (about $1''$, which corresponds to about 6.5 kpc at the redshift of the galaxies), we were unable to characterize the dynamics of the gas inside the galaxies. However, half of the galaxies showed a tentative velocity gradient in their [C II] emission^{13,14}, consistent with the [C II]

¹Max-Planck-Institut für Astronomie, Heidelberg, Germany. ²Department of Astronomy and Astrophysics, UCO/Lick Observatory, University of California, Santa Cruz, CA, USA. ³Kavli Institute for the Physics and Mathematics of the Universe (Kavli IPMU), The University of Tokyo, Kashiwa, Japan. ⁴National Centre for Radio Astrophysics, Tata Institute of Fundamental Research, Pune University, Pune, India. ⁵Space Telescope Science Institute, Baltimore, MD, USA. ⁶Department of Physics and Astronomy, Johns Hopkins University, Baltimore, MD, USA. ✉e-mail: neeleman@mpia.de

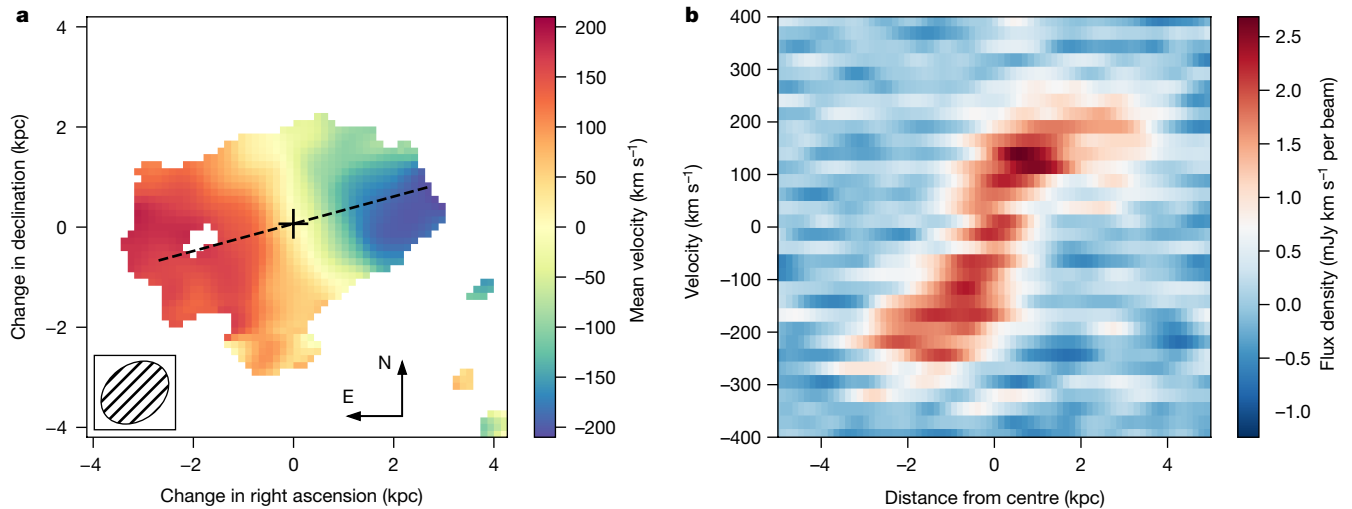


Fig. 1 | Mean velocity field and p - v diagram for DLA0817g. **a**, Mean velocity field relative to the systemic redshift of the [C II] emission ($z = 4.2603$). The kinematic centre of the [C II] emission, as determined from modelling the emission (see Methods), is shown by a black plus sign. The dotted black line

marks the major axis of the galaxy. The axes give relative physical (proper) distances at the redshift of the [C II] emission. The inset shows the synthesized beam of the observations. **b**, The p - v diagram along the major axis of the galaxy. Distances are measured from the kinematic centre of the galaxy.

arising from a rotating disk. To explore the origin of this [C II] emission, we carried out high-resolution (about $0.19''$, corresponding to approximately 1.3 kpc at the redshift of the galaxy) ALMA observations of the

[C II] and dust continuum emission from the brightest [C II]-emitting galaxy of this sample, DLA0817g, which is associated with a $z = 4.26$ absorber towards QSO J081740.52+135134.5.

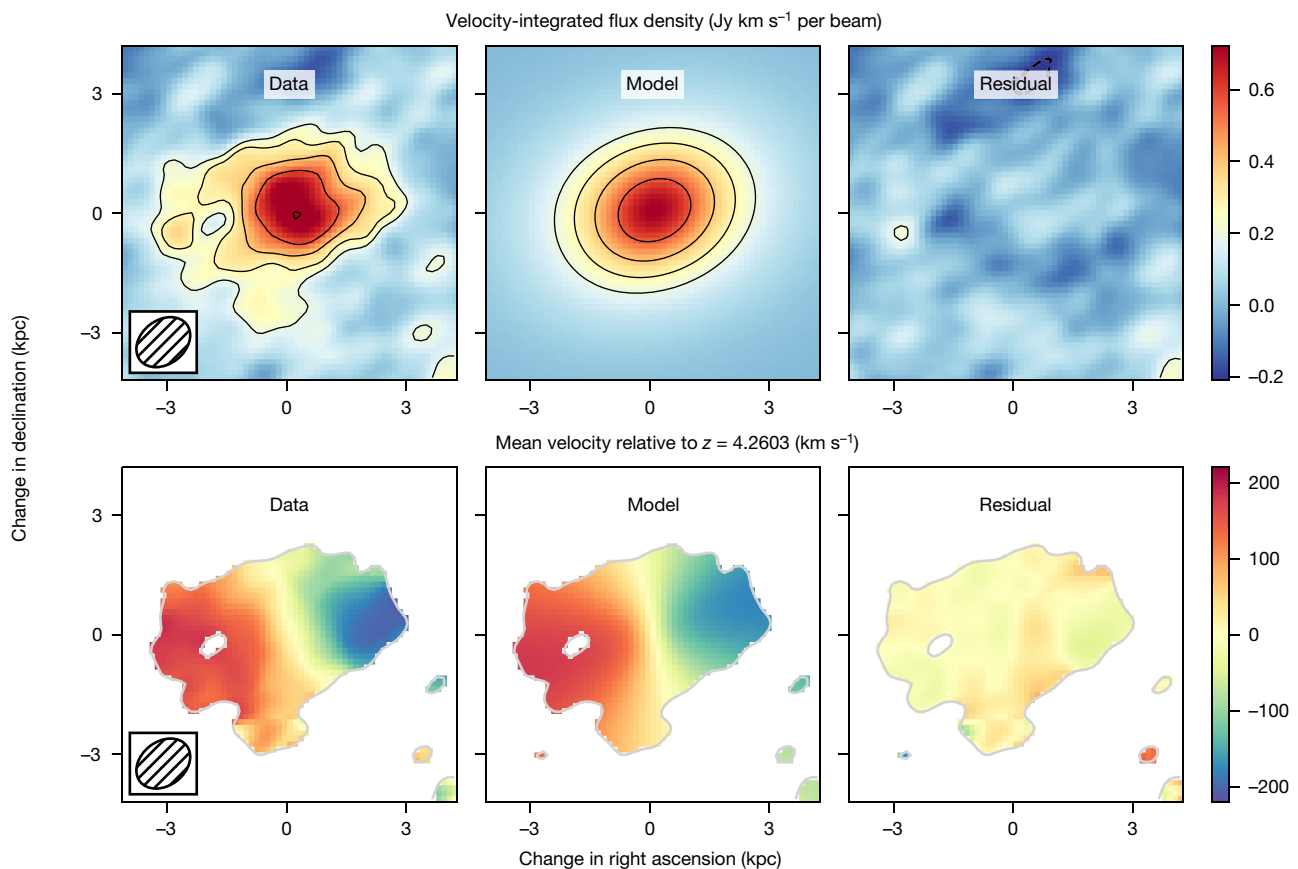


Fig. 2 | Comparison between the data and the model for DLA0817g. The top row shows the velocity-integrated [C II] flux density for the data (left panel), the constant rotational velocity model (middle panel) and the residual after subtracting the model from the data (right panel). The outer contour is at 3σ , where $\sigma = 0.0656 \text{ Jy km s}^{-1}$ is the standard deviation of the noise in the observations, with contours increasing in powers of $\sqrt{2}$. No negative contours

at the same levels are observed in the image. The synthesized beam of the observations is shown in the bottom left corner of the leftmost plot. The bottom row shows the mean velocity of the [C II] emission, for the data (left panel), the model (middle) and the residuals (right). Velocities are relative to the systemic velocity of the [C II] emission, corresponding to $z = 4.2603$.

The [C II] emission from DLA0817g is resolved in the imaging at a resolution of $0.19''$ (1.3 kpc), and has an extent of $0.63 \pm 0.07''$ (4.2 ± 0.5 kpc), along the major axis of the galaxy (see Methods). Within this extent, the velocity gradient of the [C II] line remains smooth even at $0.19''$ resolution (Fig. 1a), indicating that the [C II] emission arises from a rotationally supported disk. This is further corroborated by the position–velocity (p – v) diagram of DLA0817g (Fig. 1b), which is a slice of the spectral cube, along the galaxy’s major axis. The p – v diagram shows a flattening of the velocity curve at about 1.8 kpc, indicating that the gas has reached a constant rotational velocity. This characteristic S-shape is the proto-typical signature of a rotating disk¹⁵.

Using a custom, Python-based Markov chain Monte Carlo code, we have fitted the observations with a rotating disk model (Methods). The best-fit model has a position angle of $105.2^\circ \pm 1.7^\circ$, an inclination angle of $42_{-8}^{+3^\circ}$ and an inclination-corrected rotational velocity of 272_{-13}^{+52} km s^{−1}. The residuals after subtracting the disk model are below 3σ significance, where σ is the standard deviation of the noise, and show little velocity offset, indicating that the bulk of the [C II] emission can be modelled as a rotating disk (Fig. 2). With this position angle and inclination, we reconstruct the rotation curve of DLA0817g (Fig. 3) through two different approaches (peak velocity and mean velocity; see Methods). The rotation curve is flat beyond a radius of about 1.8 kpc, with a mean rotational velocity that is consistent with the value from the kinematic modelling. The observed decrease in the rotation curve below this radius is due to resolution effects (also known as beam-smearing); the rotating disk model, which has a constant velocity, shows the same decrease when convolved to the resolution of the data. Combining the rotational velocity estimate with the maximum extent of the [C II] emission yields a dynamical mass estimate of $(7.2 \pm 2.3) \times 10^{10} M_\odot$ (Methods).

Having established the disk origin of the [C II] emission, we can provide an estimate of the rotational support and the stability of the gas against axisymmetric perturbations. The ratio of rotational velocity to velocity dispersion (v_{rot}/σ_v) provides a measure of the rotational support of a galaxy, with ratios greater than 3 indicating a galaxy that is rotation-dominated¹⁶. For DLA0817g, the estimates of both the rotational velocity and the velocity dispersion come from the kinematic modelling of the [C II] emission line. These estimates are consistent with measurements of the rotational velocity and velocity dispersion away from the centre of emission where beam-smearing is less severe (Methods). Our estimate for v_{rot}/σ_v is $3.4_{-0.3}^{+1.1}$, consistent with DLA0817g being a rotation-dominated system. In addition to v_{rot}/σ_v , the Toomre- Q parameter of a disk provides a measure of the stability of the disk against gravitational fragmentation, where values of order unity indicate stable disks^{17,18}. For DLA0817g, we obtain a disk-averaged Toomre- Q parameter of 0.96 ± 0.30 , close to the stability limit, which is consistent with predictions from theory¹⁶. However, local values of Q within the disk can fall well below unity, resulting in unstable regions that should collapse to form dense gas and then stars¹⁹. Such dense gas is expected to show weaker [C II] emission, as most of the carbon is locked up in CO (ref. ²⁰), and the increase in dust will attenuate the [C II] emission²¹. The [C II] cavity in DLA0817g, about 2 kpc east of the galaxy centre (Fig. 1), may arise from such a locally unstable region.

Although [C II] emission is a good tracer of the dynamics of a galaxy²², it can originate from gas with a wide range of physical properties^{23,24}. Here we examine how the resolved [C II] observations compare to tracers of different mass constituents of the galaxy (that is, the dust, stellar and molecular gas components). The underlying far-infrared continuum emission in the ALMA observations arises from reprocessed dust, and thus traces the spatial distribution of dust²⁵. To trace the stellar properties of the galaxy, we observed DLA0817g with the Wide Field Camera 3 on the Hubble Space Telescope (HST), using the F160W filter. These observations cover the rest-frame near-ultraviolet (UV; about 300 nm) emission from the galaxy at an angular resolution of $0.31''$. Overlaying the [C II] and dust continuum contours on the HST

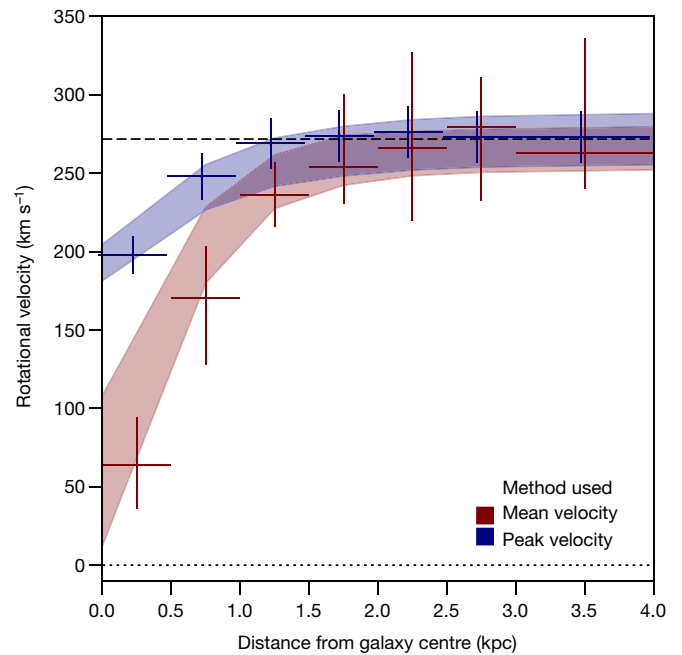


Fig. 3 | Rotation curve for DLA0817g. The rotation curve is derived from two different approaches, one approach (red curve) uses the mean velocity field to calculate the rotational velocity, and the other approach (blue curve) estimates the peak rotational velocity in the full data cube (Methods). The points show the observational data of the [C II] emission line. The horizontal error bars indicate the size of the radial bins, whereas the vertical error bars indicate the 16 to 84 percentile range of the individual measurements within the bin. The shaded region marks the 16 to 84 percentile range when the same method is applied to the rotating disk model with constant velocity. The dashed line indicates a rotational velocity of 272 km s^{-1} , as estimated from the kinematic modelling. The decrease in the model and data curves below a distance of about 1.8 kpc is due to convolution with the ALMA beam, which has a size of approximately 1.3 kpc.

image (Fig. 4b) shows that the emission centroids of all the tracers agree within the uncertainties, as do their effective radii (Methods). The dust continuum emission shows no evidence of substructure on scales of approximately 1.3 kpc, consistent with an origin in a smooth disk. The stellar emission from DLA0817g is extended with a physical size of $1.2 \pm 0.5''$ (8.1 ± 3.4 kpc) along the major axis of the galaxy at a position angle of $114 \pm 8^\circ$, within one standard deviation from the position angle obtained from the ALMA observations. The similar extent and shape of the near-UV, [C II] and dust emission suggests that the [C II] emission predominantly traces gas that is co-spatial with the stars and the dust.

To estimate the molecular content of DLA0817g, we observed the redshifted CO(2–1) transition with the JVLA. The CO rotational lines provide the best tracer of the molecular gas, as CO is the second-most abundant molecule (after molecular hydrogen, which is difficult to detect directly) in the interstellar medium of galaxies. The JVLA observations yield a detection of the CO(2–1) line at the position of DLA0817g (Fig. 4a), which results in an estimate of the molecular mass of $(8.8 \pm 2.6) \times 10^{10} \times (0.81/r_{2,1}) \times (\alpha_{\text{CO}}/3.0) M_\odot$ (Methods). This molecular mass estimate is comparable with our estimate for the dynamical mass of DLA0817g, with the caveat that the unresolved JVLA observations can only probe the total molecular mass within the beam of the JVLA observations (~ 15 kpc). However, under the assumption that most of the molecular gas is constrained within the region of the galaxy that contains stars, dust and [C II] emission²⁶, this implies that a substantial fraction of the galaxy’s mass must reside in a cold, dense gas phase.

DLA0817g was identified owing to an enriched neutral hydrogen (HI) absorber at a projected distance of $6.2''$ (42 kpc) from the galaxy¹³, with an HI column density of $2 \times 10^{21} \text{ cm}^{-2}$. In the local Universe, such

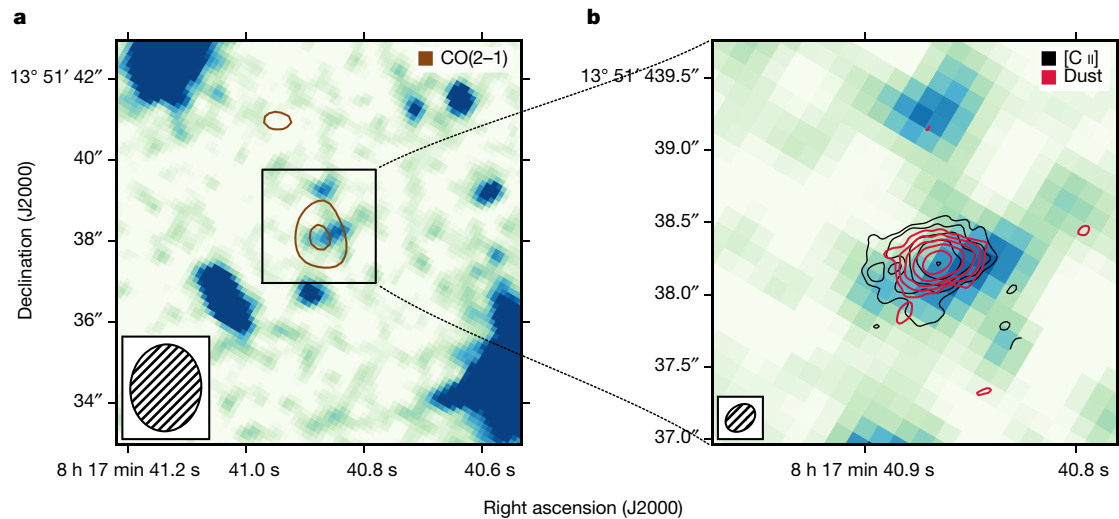


Fig. 4 | HST imaging of DLA0817g with CO, [C II] and dust continuum contours. The F160W filter on HST's Wide Field Camera 3 was used to probe the rest-frame near-UV emission of the galaxy. **a**, The field surrounding DLA0817g. The bright source in the bottom right corner is the quasar whose sightline contains the $z = 4.26$ absorber. Overlaid on this figure in contours is the CO(2-1) emission obtained with the JVLA. Contours start at 3σ , where $\sigma = 36 \mu\text{Jy}$ per beam is the standard deviation of the noise in the observation, and increase in powers

of $\sqrt{2}$. No negative contours are seen in this image at similar levels. The inset shows the synthesized beam of the JVLA observations. **b**, An enlargement of the near-UV emission from DLA0817g. Overlaid on the figure are the contours of [C II] (black) and dust continuum (red) emission obtained with ALMA. The synthesized beam of the ALMA observations is shown in the inset. Contours start at 3σ and increase in powers of $\sqrt{2}$, where $\sigma = 0.0656 \text{ Jy km s}^{-1}$ and $27 \mu\text{Jy}$ per beam for the [C II] and dust continuum observations, respectively.

high H I column densities are found only within the disks of galaxies²⁷. It is, however, unlikely that the absorption arises from an extension of the galaxy disk that we have imaged in [C II], as the implied disk radius of $\geq 42 \text{ kpc}$ is far larger than values predicted by numerical simulations⁸. The absorbing gas is more likely to arise in a gas-rich clump inside an extended H I reservoir that has been previously enriched by DLA0817g and is co-rotating with the disk¹³. This further disfavours 'hot-mode' accretion as the primary mass accretion scenario, as such high column densities of cold neutral gas are not expected so far away from the central galaxy in this scenario.

The properties of DLA0817g and its associated absorber are typical for the full sample of metal-enriched, absorption-selected galaxies at these redshifts¹⁴. Together with the selection method, through absorption, this suggests that such systems are common among normal star-forming galaxies at these redshifts. These observations therefore disfavour 'hot-mode' accretion as the primary mass accretion method for these galaxies and support the existence of cold, rotationally supported disk galaxies when the Universe was about 10% of its current age.

Online content

Any methods, additional references, Nature Research reporting summaries, source data, extended data, supplementary information, acknowledgements, peer review information; details of author contributions and competing interests; and statements of data and code availability are available at <https://doi.org/10.1038/s41586-020-2276-y>.

- Rees, M. J. & Ostriker, J. P. Cooling, dynamics and fragmentation of massive gas clouds: clues to the masses and radii of galaxies and clusters. *Mon. Not. R. Astron. Soc.* **179**, 541–559 (1977).
- Fall, S. M. & Efstathiou, G. Formation and rotation of disc galaxies with haloes. *Mon. Not. R. Astron. Soc.* **193**, 189–206 (1980).
- Kereš, D., Katz, N., Weinberg, D. H. & Davé, R. How do galaxies get their gas? *Mon. Not. R. Astron. Soc.* **363**, 2–28 (2005).
- Dekel, A. et al. Cold streams in early massive hot haloes as the main mode of galaxy formation. *Nature* **457**, 451–454 (2009).
- Hodge, J. A. et al. Evidence for a clumpy, rotating gas disk in a submillimeter galaxy at $z = 4$. *Astrophys. J.* **760**, 11 (2012).
- Smit, R. et al. Rotation in [C II]-emitting gas in two galaxies at a redshift of 6.8. *Nature* **553**, 178–181 (2018).

- Grand, R. J. J. et al. The Auriga Project: the properties and formation mechanisms of disc galaxies across cosmic time. *Mon. Not. R. Astron. Soc.* **467**, 179–207 (2017).
- Pillepich, A. et al. First results from the TNG50 simulation: the evolution of stellar and gaseous disks across cosmic time. *Mon. Not. R. Astron. Soc.* **490**, 3196–3233 (2019).
- Ade, P. A. R. et al. (Planck Collaboration) Planck 2015 results. XIII. Cosmological parameters. *Astron. Astrophys.* **594**, A13 (2016).
- Förster Schreiber, N. M. et al. The SINS Survey: SINFONI integral field spectroscopy of $z \sim 2$ star-forming galaxies. *Astrophys. J.* **706**, 1364–1428 (2009).
- Price, S. H. et al. The MOSDEF Survey: dynamical and baryonic masses and kinematic structures of star-forming galaxies at $1.4 \leq z \leq 2.6$. *Astrophys. J.* **819**, 80 (2016).
- Genzel, R. et al. Strongly baryon-dominated disk galaxies at the peak of galaxy formation ten billion years ago. *Nature* **543**, 397–401 (2017).
- Neeleman, M. et al. [C II] 158- μm emission from the host galaxies of damped Lyman-alpha systems. *Science* **355**, 1285–1288 (2017).
- Neeleman, M., Kanekar, N., Prochaska, J. X., Rafelski, M. A. & Carilli, C. L. [C II] 158- μm emission from $z \sim 4$ H I absorption-selected galaxies. *Astrophys. J.* **870**, L19 (2019).
- de Blok, W. J. G. et al. High-resolution rotation curves and galaxy mass models from THINGS. *Astron. J.* **136**, 2648–2719 (2008).
- Burkert, A. et al. High-redshift star-forming galaxies: angular momentum and baryon fraction, turbulent pressure effects, and the origin of turbulence. *Astrophys. J.* **725**, 2324–2332 (2010).
- Toomre, A. On the gravitational stability of a disk of stars. *Astrophys. J.* **139**, 1217–1238 (1964).
- Goldreich, P. & Lynden-Bell, D. I. Gravitational stability of uniformly rotating disks. *Mon. Not. R. Astron. Soc.* **130**, 97 (1965).
- Elmegreen, B. G., Bournaud, F. & Elmegreen, D. M. Bulge formation by the coalescence of giant clumps in primordial disk galaxies. *Astrophys. J.* **688**, 67–77 (2008).
- Beuther, H. et al. Carbon in different phases ([C II], [C I], and CO) in infrared dark clouds: cloud formation signatures and carbon gas fractions. *Astron. Astrophys.* **571**, A53 (2014).
- Riechers, D. A. et al. A dust-obscured massive maximum-starburst galaxy at a redshift of 6.34. *Nature* **496**, 329–333 (2013).
- de Blok, W. J. G. et al. Comparing [C II], H I, and CO dynamics of nearby galaxies. *Astron. J.* **152**, 51 (2016).
- Croxall, K. V. et al. The origins of [C II] emission in local star-forming galaxies. *Astrophys. J.* **845**, 96 (2017).
- Cormier, D. et al. The Herschel Dwarf Galaxy Survey. II. Physical conditions, origin of [C II] emission, and porosity of the multiphase low-metallicity ISM. *Astron. Astrophys.* **626**, A23 (2019).
- Carilli, C. L. & Walter, F. Cool gas in high-redshift galaxies. *Annu. Rev. Astron. Astrophys.* **51**, 105–161 (2013).
- Gullberg, B. et al. The dust and [C II] morphologies of redshift ~ 4.5 submillimeter galaxies at $\sim 200 \text{ pc}$ resolution: the absence of large clumps in the interstellar medium at high-redshift. *Astrophys. J.* **859**, 12 (2018).
- Jones, M. G., Haynes, M. P., Giovanelli, R. & Moorman, C. The ALFALFA H I mass function: a dichotomy in the low-mass slope and a locally suppressed 'knee' mass. *Mon. Not. R. Astron. Soc.* **477**, 2–17 (2018).

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

© The Author(s), under exclusive licence to Springer Nature Limited 2020

Methods

Cosmology

Throughout this paper we use a flat Λ Cold Dark Matter cosmology, defined by the parameters $\Omega_\Lambda = 0.7$, $\Omega_m = 0.3$ and $H_0 = 70 \text{ km s}^{-1} \text{ Mpc}^{-1}$. Adopting this concordance cosmology aids comparisons with previous results in the literature.

ALMA observations and data reduction

ALMA observed the galaxy associated with the absorber towards QSO J081740.52+135134.5, DLA0817g, in two executions on 2018 January 6 and 2018 January 21 (all dates are given in Universal Time) for a total on-source time of 100 min (programme ID 2017.1.01052.S, Principal Investigator (PI): Neeleman). One of the four 1.875-GHz ALMA receiver bands was centred at 361.312 GHz, the redshifted frequency of the [C II] 158- μm ([C II]) emission line, with 240 channels, that is, a resolution of 7.8125 MHz per channel. The remaining three 1.875-GHz bands were set up for continuum observations at neighbouring frequencies. The blazar J0854+2006 was used as a flux and bandpass calibrator, and the blazar J0750+1231 as the phase calibrator. The nominal antenna configuration was C43-5, with a maximum baseline length of approximately 2.5 km. A total of 43 and 44 antennas were used for the two executions, respectively.

The raw data were calibrated using the ALMA calibration pipeline, which is part of the Common Astronomy Software Application (CASA) package²⁸. The calibrated visibility data set was then re-examined within CASA, where minor additional flagging was performed. From the line-free channels, we generated two continuum images by using the task *tclean*, one using natural weighting, and the other using Briggs weighting with a robust parameter of 0.5. The resultant synthesized beams are $0.23'' \times 0.17''$ and $0.16'' \times 0.12''$ with root-mean-square (r.m.s.) sensitivities of 23 μJy per beam and 27 μJy per beam, respectively. To generate the [C II] emission line cube, we subtracted the continuum emission using the task *uvsub* and then subtracted any remaining residual continuum emission with *uvcontsub*. As with the continuum image, we made two different image cubes using *tclean* for the two different weighting schemes resulting in r.m.s. sensitivities of 0.35 mJy per beam per 25 km s^{-1} for natural weighting and 0.41 mJy per beam per 25 km s^{-1} for Briggs weighting with a robust parameter of 0.5. In this paper, we opted to use only the natural weighting scheme, as the better sensitivity provided slightly better constraints than the higher resolution of the Briggs weighting scheme. However, none of the results presented here are affected by the choice of this weighting scheme.

JVLA observations and data reduction

To search for redshifted CO(2–1) emission from DLA0817g, we observed the galaxy using the JVLA in four separate observing runs between 2017 March 03 and 2017 April 10, for a total on-source integration time of 9 h (programme ID 17-279, PI: Neeleman). One of the two JVLA intermediate frequency bands was centred on the redshifted CO(2–1) line at 43.828 GHz, with the three central sub-bands having a resolution of 250 kHz. The remaining sub-bands were set up at a coarser resolution of 1 MHz. The most compact configuration, D, was used, resulting in a synthesized beam of $2.1'' \times 1.8''$. We performed the calibration of the raw data using standard routines within CASA. The calibrated individual runs were combined using the task *concat* and then a continuum image was made from the line-free channels, using the task *tclean*. No continuum emission was detected within the primary beam of the observations, which covers both the galaxy, DLA0817g, at $z = 4.2603$, and the quasar, SDSS J081740.52+135134.5, at $z = 4.398$. Two emission line cubes were created of the redshifted CO(2–1) line from DLA0817g with *tclean* and natural weighting, one with a channel width of 50 km s^{-1} , and the other with a width of 550 km s^{-1} , with r.m.s. sensitivities of 120 μJy per beam and 36 μJy per beam, respectively. The latter cube was used to measure the CO(2–1) line flux density, and create the contours in Fig. 4.

HST observations, data reduction and analysis

HST observations of DLA0817g (programme ID 15410, PI: Neeleman) were obtained on 2018 May 07 and consist of one orbit of four 653-s exposures with the Wide Field Camera 3, using the F160W filter. This filter covers the rest-frame near-UV stellar light from the galaxy. We created the image mosaic using AstroDrizzle and aligned the astrometry to the GAIA DR2 astrometry²⁹ using TweakReg, resulting in an absolute astrometric uncertainty of approximately 0.006''. The effective spatial resolution of the HST observations is 0.31'', as determined from a Gaussian fit to the point spread function of the quasar.

We detect rest-frame near-UV emission from DLA0817g at 7σ significance within the isophotal area of the galaxy (22 pixels). To determine the photometry and basic shape of the stellar light from DLA0817g, we used Source Extractor (v.2.5.0³⁰). The galaxy is extended with an ellipticity of 0.6 and a position angle of $114^\circ \pm 8^\circ$, fully consistent with the position angle determined from the kinematic analysis of the [C II] line emission (see kinematic analysis section below). We measure the total flux using Source Extractor's *flux_auto*, which provides the flux within an elliptical aperture with the Kron radius³¹. This yields a total AB magnitude of $M_{\text{AB}} = 25.1 \pm 0.2$ and a Kron radius of 0.9''.

Far-infrared continuum, [C II] line and CO(2–1) line luminosities

The rest-frame 160- μm continuum flux density of DLA0817g is detected and resolved by the ALMA observations with a flux density within a $1.5''$ radius centred on the galaxy of $1.28 \pm 0.15 \text{ mJy}$. The size of the emission is $(0.41 \pm 0.05'' \times 0.24 \pm 0.03'')$, as determined from a Gaussian fit to the data. These measurements are consistent with the values obtained from the lower-resolution measurement¹³, indicating that no emission is resolved out by the higher-resolution ALMA observations. From this single far-infrared measurement, we can estimate the total far-infrared luminosity (L_{TIR} , defined as the integrated luminosity between 8 μm and 1,000 μm), assuming that the emission has a modified blackbody spectrum, and with the caveat that neither the dust temperature (T_d) nor the power law spectral index (β) for the dust emissivity is constrained by this single measurement. Assuming fiducial values of $T_d = 35 \text{ K}$ and $\beta = 1.6$, with possible ranges of $25 \text{ K} < T_d < 45 \text{ K}$, and $1.2 < \beta < 2.0$, gives a total far-infrared luminosity estimate of $L_{\text{TIR}} = 1.2^{+2.3}_{-0.7} \times 10^{12} L_\odot$ (L_\odot , luminosity of the Sun), consistent with previous estimates¹³. All observational measurements are listed in Extended Data Table 1.

To estimate the total flux from the [C II] emission, we measure the flux density within a radius of $1.5''$ centred on DLA0817g for each channel. The resultant spectrum is shown in Extended Data Fig. 1. To estimate the velocity-integrated line flux density, we integrated the spectrum over the channels showing [C II] emission (marked by the horizontal bar in Extended Data Fig. 1). The total velocity-integrated [C II] flux density from DLA0817g is $5.8 \pm 0.4 \text{ Jy km s}^{-1}$ resulting in a [C II] luminosity of $L_{[\text{C II}]} = (3.26 \pm 0.22) \times 10^9 L_\odot$. This value is consistent with that obtained from the lower-resolution data¹³, indicating that the [C II] emission is not resolved out in the higher-resolution observations. To estimate the extent of the emission we fit a Gaussian profile to the velocity-integrated map of the [C II] emission, resulting in a size of $(0.63 \pm 0.07'' \times 0.43 \pm 0.05'')$.

An emission line is detected in the JVLA data cube at 4.7σ significance, with a velocity-integrated flux density of $94 \pm 20 \text{ mJy km s}^{-1}$. Both the spatial location and the velocity of this spectral feature are in excellent agreement with the position and velocity of the [C II] emission line from DLA0817g, strengthening the identification of the feature as the CO(2–1) line from the galaxy. Taking into account that the observations are taken against the cosmic microwave background (CMB) results in a correction factor of $1/(1 - B_\nu[T_{\text{CMB}}]/B_\nu[T_{\text{exc}}])$, where T_{CMB} is the temperature of the CMB at the redshift of DLA0817g, T_{exc} is the excitation temperature of the CO(2–1) line, and $B_\nu[T]$ is the black body intensity at temperature T and the frequency of the CO(2–1) line³².

Article

As the excitation temperature is not known, we take a range of values between 25 K and 45 K, resulting in a correction factor of 1.5 ± 0.3 and an intrinsic velocity-integrated CO(2–1) flux density of $0.14 \pm 0.04 \text{ Jy km s}^{-1}$. Converting the velocity-integrated flux density into a line luminosity with standard equations²⁵ gives a line luminosity of $L'_{\text{CO}(2-1)} = (2.4 \pm 0.7) \times 10^{10} \text{ K km s}^{-1} \text{ pc}^2$ or $L_{\text{CO}(2-1)} = (9.6 \pm 2.8) \times 10^6 L_{\odot}$.

Estimates of star-formation rates

We can convert both the far-infrared dust continuum measurement and [C II] luminosity into an SFR estimate. Dust continuum emission is expected to be a good tracer of the SFR as the far-infrared emission arises from dust that has been heated by stellar UV photons. To estimate the SFR from the far-infrared continuum measurement, we use the observationally determined calibration between SFR and 160- μm continuum³³, with the caveat that a fraction of the emission might arise from a population of older stars not associated with the current SFR³⁴. However, this fraction is small for galaxies with comparable dust continuum³³. The measured star formation rate from the 160- μm continuum is $\text{SFR}_{160\mu\text{m}} = (118 \pm 14) M_{\odot} \text{ yr}^{-1}$. This value is in rough agreement with the SFR as determined from converting the total infrared luminosity (TIR) into an SFR by using the scaling relationship given for local galaxies³⁵: $\text{SFR}_{\text{TIR}} = 177^{+344}_{-103} M_{\odot} \text{ yr}^{-1}$. Finally, observations at both low and high redshifts have shown a relationship between SFR and [C II] luminosity^{36,37}. Using this relationship gives an SFR of $\text{SFR}_{[\text{C II}]} = (420 \pm 260) M_{\odot} \text{ yr}^{-1}$. This is higher than the previous estimates, but in agreement with published results that show that DLA0817g sits slightly above this locally derived relationship¹³.

Using the rest-frame far-UV flux measurement, we can also provide an estimate of the SFR that is not obscured by dust. By applying the scaling relationship for local galaxies and assuming a Kroupa IMF³⁵, the 1.6- μm measurement of DLA0817g corresponds to $\text{SFR}_{1.6\mu\text{m}} = (16 \pm 3) M_{\odot} \text{ yr}^{-1}$. The large discrepancy between the dust-obscured SFR estimate and the far-UV SFR estimate indicates that DLA0817g has a significant amount of dust, which obscures the UV radiation.

Estimates of molecular gas mass

Several estimators have been used in the literature to estimate the molecular gas mass of high-redshift galaxies. The CO(2–1) luminosity can be converted into a molecular gas mass estimate by assuming two quantities: the luminosity ratio between the CO(2–1) and CO(1–0) emission lines, $r_{2,1}$, and the conversion factor between the CO(1–0) luminosity and the molecular gas mass, α_{CO} . If the CO emission lines were thermalized, then $r_{2,1}$ would be equal to 1. However, this is unlikely to be the case, because star-forming galaxies at lower redshifts are slightly sub-thermally populated³⁸. We therefore assume $r_{2,1} = 0.81$, as determined from these galaxies. The other parameter, α_{CO} , is less well-determined and can vary between values of about $1 M_{\odot} (\text{K km s}^{-1} \text{ pc}^2)^{-1}$ for quasars and starburst galaxies, and about $(3.0\text{--}4.3) M_{\odot} (\text{K km s}^{-1} \text{ pc}^2)^{-1}$ for the Milky Way and other disk galaxies, to $\gg 10 M_{\odot} (\text{K km s}^{-1} \text{ pc}^2)^{-1}$ for low-metallicity galaxies³⁹. As DLA0817g has physical properties similar to those of regular star-forming galaxies, and not the extreme conditions present in rapidly star-forming quasars and starburst galaxies nor the low SFR and metallicity of dwarf galaxies, we assume a conservative α_{CO} of $3.0 M_{\odot} (\text{K km s}^{-1} \text{ pc}^2)^{-1}$ as was measured for colour-selected galaxies at high redshift⁴⁰. The resultant molecular gas mass estimate is $M_{\text{mol,CO}} = (8.8 \pm 2.6) \times 10^{10} \times (0.81/r_{2,1}) \times (\alpha_{\text{CO}}/3.0) M_{\odot}$.

The molecular gas mass of a galaxy may also be estimated via two other, more indirect, methods, using conversion factors from the far-infrared continuum luminosity⁴¹ or the [C II] line luminosity⁴² to the molecular gas mass. The mass estimates from these two methods are $M_{\text{mol,FIR}} = (5.7 \pm 0.7) \times 10^{10} \times (6.7 \times 10^{19}/\alpha_{850\mu\text{m}}) M_{\odot}$ and $M_{\text{mol,[C II]}} = (9.8 \pm 0.6) \times 10^{10} \times (\alpha_{[\text{C II}]})/30 M_{\odot}$, respectively. The results from both methods are in agreement with the estimate of molecular gas mass based on the CO(2–1) emission line.

For all molecular mass estimates, we only report the observational uncertainties. The uncertainties due to the conversion factors are considered in the measurement by reporting the molecular mass as a function of the conversion factor. For both the molecular gas mass estimates based on the far-infrared continuum and that based on the [C II] line luminosity, these uncertainties have not been studied in detail as a function of galaxy properties, but are presumed to be larger than the observational uncertainties; for example, for the [C II] conversion factor, $\alpha_{[\text{C II}]}$, the scatter is at least a factor of 2 (ref. ⁴²), while a similar scatter of $\sim 50\%$ is observed in the far-infrared continuum conversion factor, $\alpha_{850\mu\text{m}}$ (ref. ⁴¹). The conversion factor from the CO(2–1) emission line to molecular mass is a function of the galaxy's star formation and metallicity³⁹, and is better calibrated than the other measurements. We therefore report the estimate of molecular gas mass based on the CO(2–1) emission line.

Kinematic analysis of the [C II] emission line

To model the dynamics of the [C II] emission line, we have fitted the observed data cube using a custom Python programme that generates a model cube from a user-defined emission model. The programme convolves the model cube with the observed beam and then minimizes the residuals between the convolved model and the observed data cube using a Markov chain Monte Carlo approach^{43–45}. This programme also yields estimates of the uncertainty on each of the parameters of the model.

For DLA0817g, we assume an emission model in which the [C II] emission arises from a thin disk in which the emission exponentially declines: $I(R) = I_0 e^{-R/R_d}$. The velocity curve of the [C II] emission is assumed either to be constant as a function of radius, $v(R) = v_{\text{rot}}$, or to increase with radius via an arctangent profile, $v(R) = (2v_{\text{rot}}/\pi) \arctan(R/R_v)$. In both models, we assume a constant velocity dispersion across the disk, $\sigma(R) = \sigma_v$. Together with the three spatial coordinates (x_0, y_0, z_0), and the inclination (i) and position angle (α), these nine or ten parameters uniquely determine the [C II] line emission in the model cube.

The results of the Markov chain Monte Carlo analysis are given in Extended Data Table 2. No significant differences are found between the two different models, and therefore we opt for the model with the fewest free parameters, the constant rotational velocity model. The top-left and top-middle panels of Fig. 2 show the velocity-integrated [C II] flux density of DLA0817g for the observed data and the model with constant rotational velocity. The top-right panel shows the residual between the model and the data, which lacks any features of $\geq 3\sigma$ significance. The bottom panels show the velocity field of the model and observations. For completeness, we show the channel maps of DLA0817g and of the best-fit model (Extended Data Figs. 2 and 3) and the position–velocity diagrams along the major and minor axes (Extended Data Fig. 4). Little excess emission (with $\geq 3\sigma$ significance) is seen in the channel maps of the residual, suggesting that the simple exponential thin-disk, constant rotational velocity model is an accurate representation of the bulk of the [C II] emission.

In the above models, we assume the [C II] emission arises from an infinitely thin disk. This is a simplified assumption, as disks at high redshifts are expected to be turbulent and thus thicker than local analogues⁴⁶. The high velocity dispersion in DLA0817g, compared with measurements at $z \approx 2$ from H α emission^{10,47} is a further indication that the gas disk is likely to be turbulent. To explore how the thin-disk approximation affects the kinematic analysis, we repeated the modelling for models where the [C II] distribution is distributed in a thick disk with an exponential scale height whose scale radius is varied from 0.15 to 1.0 times the scale length in the radial direction. These models have systematically lower inclinations by an average of 7° , and thus higher rotational velocities by $\sim 50 \text{ km s}^{-1}$. Furthermore, the velocity dispersion estimates are lower by 10 km s^{-1} , as some of the dispersion arises from line-of-sight motion in the thick disk. Combining both results would increase v_{rot}/σ_v by 30%. As no information is available about the

thickness of the disk, we take a conservative approach and report values based on the thin-disk approximation, where the uncertainties include the spread due to the thickness of the disk. As a consistency check, we also fitted DLA0817g using the ‘tilted-ring’ fitting programme ^{3D}Barolo⁴⁸, which yields a rotational velocity of 279 km s⁻¹ and a velocity dispersion of 72 km s⁻¹, consistent with our values.

Estimates of dynamical mass

To estimate the dynamical mass, we assume that the gas is rotationally supported. With this assumption, the dynamical mass (M_{dyn}) of a system (in M_\odot) within a radius R (in kpc) is given by: $M_{\text{dyn}}(R) = 2.32 \times 10^5 \nu_{\text{rot}}(R)^2 R$, where $\nu_{\text{rot}}(R)$ is the circular velocity (in km s⁻¹) of the galaxy at the radius R (ref. ⁴⁹). We note that in deriving this equation, the mass distribution is implicitly assumed to be spherical. For an exponential thin-disk mass distribution, this underestimates the mass by up to 30%, depending on the radius at which the rotational velocity is measured⁵⁰.

Our kinematic analysis provides an estimate for the constant rotational velocity, $\nu_{\text{rot}} = 272_{-13}^{+52}$ km s⁻¹. We can compare this rotational velocity estimate to estimates of the rotational velocity from other estimators used in lower resolution data, such as the full-width at half-maximum (FWHM) of the integrated [C II] emission line. By fitting a double Gaussian profile to the integrated [C II] spectrum (Extended Data Fig. 1), we measure an FWHM_[C II] of 400 ± 40 km s⁻¹. If we assume the emission arises from ordered rotation and can be described by a sharp double Gaussian profile, then $\nu_{\text{rot}} = 0.5 \times \text{FWHM}_{[\text{C II}]} / \sin i$ (refs. ^{49,51}). With the inclination estimate of $i = 42_{-8}^{+3}$ °, we get a rotational velocity of 300 ± 50 km s⁻¹. This estimate is consistent with the rotational velocity estimate derived from our kinematic modelling, but we note that estimating the rotational velocity from the integrated line spectrum requires both an accurate inclination and a resolved spectral line profile.

To measure a dynamical mass for a galaxy, we must define a radius at which to measure this mass. Most observations of high-redshift galaxies barely resolve the emission, and the radius chosen is either the maximum extent of the emission⁴⁹ or the major semi-axis of the two-dimensional Gaussian fit to the [C II] emission⁵². In our resolved observations, we will assume the extent of the galaxy to be three times the exponential scale length of the modelled emission. This extent ($R = 4.2$ kpc $= 0.6''$) corresponds to a region that emits 80% of the total [C II] flux density of the galaxy. It is also similar to the maximum extent of the [C II] line emission that yields a reliable rotational velocity measurement (Fig. 1). Within this region, we measure a dynamical mass of $M_{\text{dyn}} = (7.2 \pm 2.3) \times 10^{10} M_\odot$, where the uncertainty includes, in quadrature, a 30% uncertainty from the assumption of a spherical mass distribution.

Mean velocity field, velocity curves and light profiles

To estimate the mean velocity field for DLA0817g (Fig. 1), we fit a Gaussian profile to the spectrum of each spatial position where the velocity-integrated flux is detected at $>3\sigma$. The mean of the Gaussian fit is the estimated mean velocity at that position. This method is more robust against flux outliers than the standard first moment map¹⁵. Using this velocity field and the de-projected distance of the pixel—determined from the inclination, galaxy centre and position angle from the kinematic modelling—yields an estimate of the rotation velocity at each pixel. Here we assume that all of the velocity is constrained within the plane of the galaxy. The resultant median and 1σ spread in data points per radial bin are shown in Fig. 3 in red, and labelled as the mean velocity method. For the second method, we take the full data cube and de-project the cube into the plane of the galaxy, again assuming that all of the velocity is constrained within the plane of the galaxy. We then take the intensity of each 3D pixel and its associated rotational velocity as the probability that the rotational velocity has that value at that position. By averaging this over all the 3D pixels within a radial bin, we get a probability distribution function of the rotational velocity

per radial bin. The rotational velocity per bin is taken to be the peak of a spline fit to the probability distribution function and is shown by the blue points in Fig. 3, labelled as the peak velocity method. In both cases, uncertainties are convolved with the uncertainties in both the inclination and position angle. The peak of a distribution is less affected by asymmetric profiles, such as those arising from beam-smearing, explaining why the peak velocity method yields better constraints closer to the centre of the galaxy.

To provide a validity check for our measurement of the velocity dispersion, we create a velocity dispersion radial profile. This profile is generated from the measurements of the width (that is, standard deviation) of the fitted Gaussian profiles, and the de-projected distance at each pixel. The resultant median and 1σ spread in data points per radial bin are shown in Extended Data Fig. 5. The horizontal dashed line indicates the value for the velocity dispersion as determined from the kinematic modelling. This figure shows that away from the kinematic centre, where beam-smearing is less severe, the observed velocity dispersion is in agreement with the velocity dispersion determined from the kinematic modelling.

The light profiles for the dust continuum, [C II] and near-UV emission are shown in Extended Data Fig. 6. The dust and near-UV continuum emission light profiles are directly measured from the continuum maps of the ALMA and HST data, respectively. For the [C II] emission, we create an integrated flux-density map over the channels highlighted in Extended Data Fig. 1 (see also Fig. 2). We do not attempt a CO(2–1) light profile, as the emission is not resolved. Before generating the light profiles, the [C II] and dust continuum are convolved with a Gaussian kernel to account for the slightly worse resolution of the near-UV data. All of the light profiles are plotted against the de-projected radius, assuming the same inclination, position angle and kinematic centre. The points are radially binned, and the measurements include systematic uncertainties from taking the same position and orientation between the different emission profiles. These measurements show that the dust continuum and the near-UV are consistent within the uncertainties with the [C II] emission.

Toomre-Q parameter

The Toomre-Q parameter is a quantity that defines how stable a system is against gravitational perturbations¹⁷. For a collisionless gas, this parameter is given by the equation: $Q = \sigma_v \kappa / \pi G \Sigma_{\text{gas}}$ (ref. ¹⁸), with $Q < 1$ corresponding to gas that is unstable. In this equation, κ is the epicyclic frequency, which, for a constant-velocity thin disk, is given by $\kappa = \sqrt{2} \nu_{\text{rot}} / R$, and Σ_{gas} is the gas surface density. If we assume that the [C II] emission line traces the gas surface density, then—together with the total molecular gas mass derived from the CO(2–1) line—we can estimate the gas surface density. The spatially resolved, beam-convolved Toomre-Q parameter distribution for DLA0817g is shown in Extended Data Fig. 7a.

Although the [C II] emission line is resolved, the limited resolution (that is, beam-smearing) affects the measurement of Q . To explore the effect of resolution on the measurement of Q , we can estimate an azimuthally averaged Q directly from the surface density profile found in the kinematic modelling, which takes into account the limited resolution of the observations. The resultant radial profile for Q is shown by the black line in Extended Data Fig. 7b. This resolution-independent measurement shows that at all radii beam-smearing causes Q to be systematically lower, because the emission is spread out over a larger region. To estimate a net Toomre-Q parameter for the whole galaxy, \bar{Q} , we average the radial profile for Q over the extent of the emission ($R = 3R_d$) according to

$$\bar{Q} = \frac{2\pi \int_0^{3R_d} Q(R) R dR}{2\pi \int_0^{3R_d} R dR} = \frac{4\sqrt{2} \sigma_v \nu_{\text{rot}} R_d}{9GM_{\text{tot}}} (e^3 - 1) \quad (1)$$

Article

Here, the last equality holds because the gas surface density profile, $\Sigma_{\text{gas}}(R)$ is a simple exponential and the integral can be solved analytically. This yields $\bar{Q} = 0.96 \pm 0.30$ for DLA0817g.

We note that we calculate the Toomre- Q parameter for only the gas. The total Q parameter should also include contributions from other galaxy constituents, in particular the stars. However, the total Q parameter is the inverse of the sum of the inverses of the individual Q parameters. As such, given that the disk is already marginally unstable from the gas contributions alone, any additional component will only make the disk more unstable. Our estimate of the Q parameter can therefore be taken as an upper limit. Another important caveat is that these measurements are averages over the size of the beam, and variations—especially in the surface density—on smaller scales could mean that certain regions within the disk are stable. On smaller scales, the gas surface density is also not likely to be traced very well by the [C II] emission, especially in the higher-density regions, where either the [C II] emission is self-absorbed²¹ or most of the carbon is in the neutral state or locked up in CO (ref.²⁰). However, these regions will be unstable, and the measured Q parameter can again be taken as an upper limit.

Data availability

The data reported in this paper are available through the ALMA archive: (<http://almascience.eso.org/aq/>) with project code 2017.1.01052.S, the JVLA archive: (<https://science.nrao.edu/facilities/vla/archive/index>) with project code 17A-279, and the HST/Mikulski Archive for Space Telescopes: (<https://archive.stsci.edu/hst/>) with project code 15410.

Code availability

All of the code used to generate the kinematic modelling is available online at (<https://github.com/mneeleman/qubefit>).

28. McMullin, J. P., Waters, B., Schiebel, D., Young, W. & Golap, K. CASA architecture and applications. In *Astronomical Data Analysis Software and Systems XVI* Vol. 376 (eds Shaw, R. A. et al.) 127 (2007).
29. Brown, A. G. A. et al. (Gaia Collaboration) Gaia Data Release 2. Summary of the contents and survey properties. *Astron. Astrophys.* **616**, A1 (2018).
30. Bertin, E. & Arnouts, S. SExtractor: software for source extraction. *Astron. Astrophys. Suppl.* **117**, 393–404 (1996).
31. Kron, R. G. Photometry of a complete sample of faint galaxies. *Astrophys. J. Suppl.* **43**, 305–325 (1980).
32. da Cunha, E. et al. On the effect of the cosmic microwave background in high-redshift (sub-)millimeter observations. *Astrophys. J.* **766**, 13 (2013).
33. Calzetti, D. et al. The calibration of monochromatic far-infrared star formation rate indicators. *Astrophys. J.* **714**, 1256–1279 (2010).
34. Lonsdale Persson, C. J. & Helou, G. On the origin of the 40–120 micron emission of galaxy disks: a comparison with H α fluxes. *Astrophys. J.* **314**, 513–524 (1987).
35. Kennicutt, R. C. & Evans, N. J. Star formation in the Milky Way and nearby galaxies. *Annu. Rev. Astron. Astrophys.* **50**, 531–608 (2012).
36. De Looze, I. et al. The applicability of far-infrared fine-structure lines as star formation rate tracers over wide ranges of metallicities and galaxy types. *Astron. Astrophys.* **568**, A62 (2014).

37. Herrera-Camus, R. et al. [C II] 158- μ m emission as a star formation tracer. *Astrophys. J.* **800**, 1 (2015).
38. Dessauges-Zavadsky, M. et al. Molecular gas content in strongly lensed $z \sim 1.5$ –3 star-forming galaxies with low infrared luminosities. *Astron. Astrophys.* **577**, A50 (2015).
39. Bolatto, A. D., Wolfire, M. & Leroy, A. K. The CO-to-H $_2$ conversion factor. *Annu. Rev. Astron. Astrophys.* **51**, 207–268 (2013).
40. Daddi, E. et al. Very high gas fractions and extended gas reservoirs in $z \sim 1.5$ disk galaxies. *Astrophys. J.* **713**, 686–707 (2010).
41. Scoville, N. et al. The evolution of interstellar medium mass probed by dust emission: ALMA observations at $z = 0.3$ –2. *Astrophys. J.* **783**, 84 (2014).
42. Zanella, A. et al. The [C II] emission as a molecular gas mass tracer in galaxies at low and high redshifts. *Mon. Not. R. Astron. Soc.* **481**, 1976–1999 (2018).
43. Bañados, E. et al. The $z = 7.54$ quasar ULAS J1342+0928 is hosted by a galaxy merger. *Astrophys. J.* **881**, L23 (2019).
44. Neeleman, M. et al. Resolved [C II] emission from $z \sim 6$ quasar host-companion galaxy pairs. *Astrophys. J.* **882**, 10 (2019).
45. Venemans, B. P. et al. 400 pc imaging of a massive quasar host galaxy at a redshift of 6.6. *Astrophys. J.* **874**, L30 (2019).
46. Bird, J. C. et al. Inside out and upside down: tracing the assembly of a simulated disk galaxy using mono-age stellar populations. *Astrophys. J.* **773**, 43 (2013).
47. Simons, R. C. et al. $z \sim 2$: an epoch of disk assembly. *Astrophys. J.* **843**, 46 (2017).
48. Di Teodoro, E. M. & Fraternali, F. ^{3D}BAROLO: a new 3D algorithm to derive rotation curves of galaxies. *Mon. Not. R. Astron. Soc.* **451**, 3021–3033 (2015).
49. Wang, R. et al. Star formation and gas kinematics of quasar host galaxies at $z \sim 6$: new insights from ALMA. *Astrophys. J.* **773**, 44 (2013).
50. Walter, F., Brinks, E., Duric, N. & Klein, U. A dynamical analysis of the H II galaxy II Zwicky 33 and its low surface brightness companion. *Astron. J.* **113**, 2031–2045 (1997).
51. Ho, L. C. The CO Tully–Fisher relation and implications for the host galaxies of high-redshift quasars. *Astrophys. J.* **669**, 821–829 (2007).
52. Decarli, R. et al. An ALMA [C II] survey of 27 quasars at $z > 5.94$. *Astrophys. J.* **854**, 97 (2018).

Acknowledgements This work would not have been possible without the insights of the late A. M. Wolfe. M.N. thanks F. Walter for discussions, and S. Simha for extracting effective radii from the HST and ALMA imaging. ALMA is a partnership of the European Southern Observatory (ESO; representing its member states), the National Science Foundation (NSF; United States) and the National Institutes of Natural Sciences (Japan), together with the National Research Council (Canada), the National Science Council and the Academia Sinica Institute of Astronomy and Astrophysics (Taiwan), and the Korean Astronomy and Space Science Institute (Republic of Korea), in cooperation with the Republic of Chile. The Joint ALMA Observatory is operated by ESO, Associated Universities Incorporated (AUI) National Radio Astronomy Observatory (NRAO) and the National Astronomical Observatory of Japan. NRAO is a facility of the NSF operated under cooperative agreement by AUI. M.N. acknowledges support from the European Research Council advanced grant 740246 (Cosmic Gas). N.K. acknowledges support from the Department of Science and Technology via a Swarnajayanti Fellowship (DST/SJF/PSA-01/2012-13), and from the Department of Atomic Energy under project 12-R&D-TFR-5.02-0700. Support for programme no. 15410 was provided by the National Aeronautics and Space Administration (NASA) through a grant from the Space Telescope Science Institute, which is operated by the Association of Universities for Research in Astronomy Incorporated, under NASA contract NAS 5-26555.

Author contributions M.N. is the PI of the observing programmes. M.N. and N.K. contributed to the analysis of the ALMA and JVLA data. M.R. reduced and analysed the HST data. All authors contributed to analysis of the data and writing and editing of the manuscript.

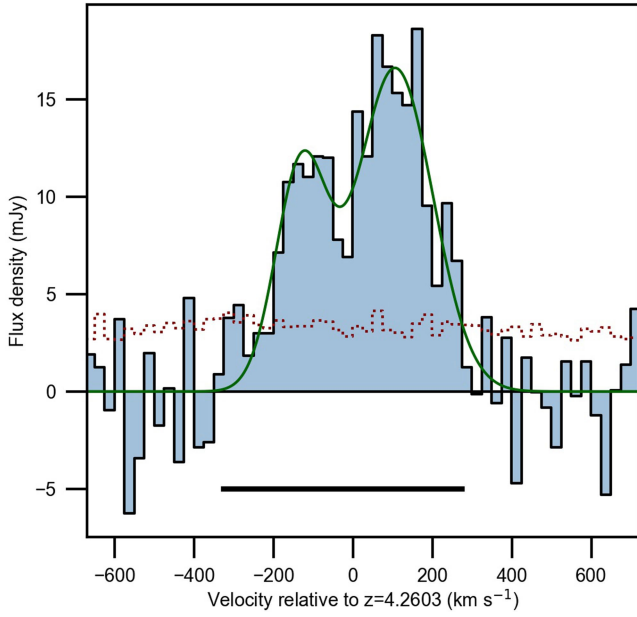
Competing interests The authors declare no competing interests.

Additional information

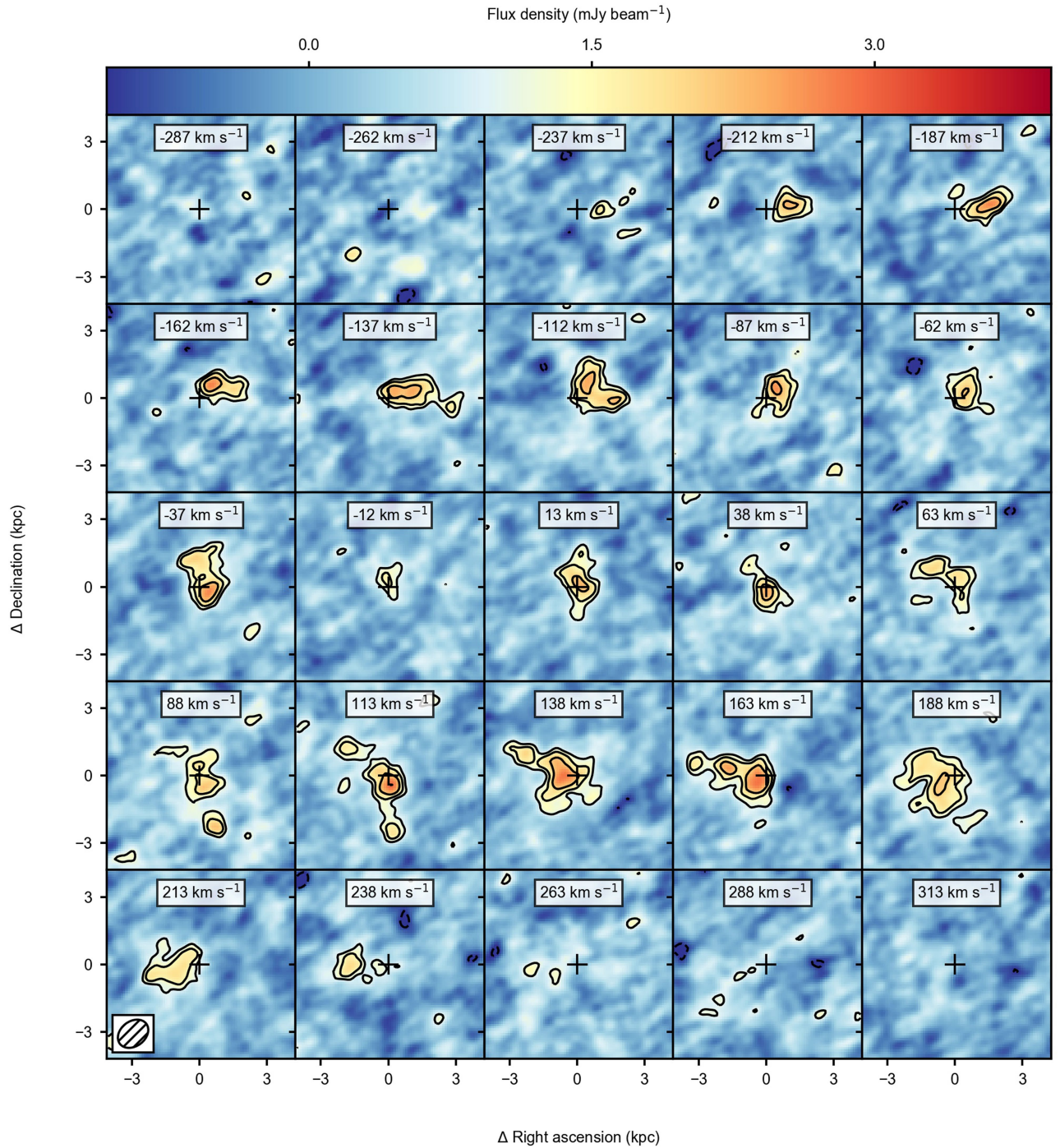
Correspondence and requests for materials should be addressed to M.N.

Peer review information Nature thanks Robert Grand and Alfred Tiley for their contribution to the peer review of this work.

Reprints and permissions information is available at <http://www.nature.com/reprints>.

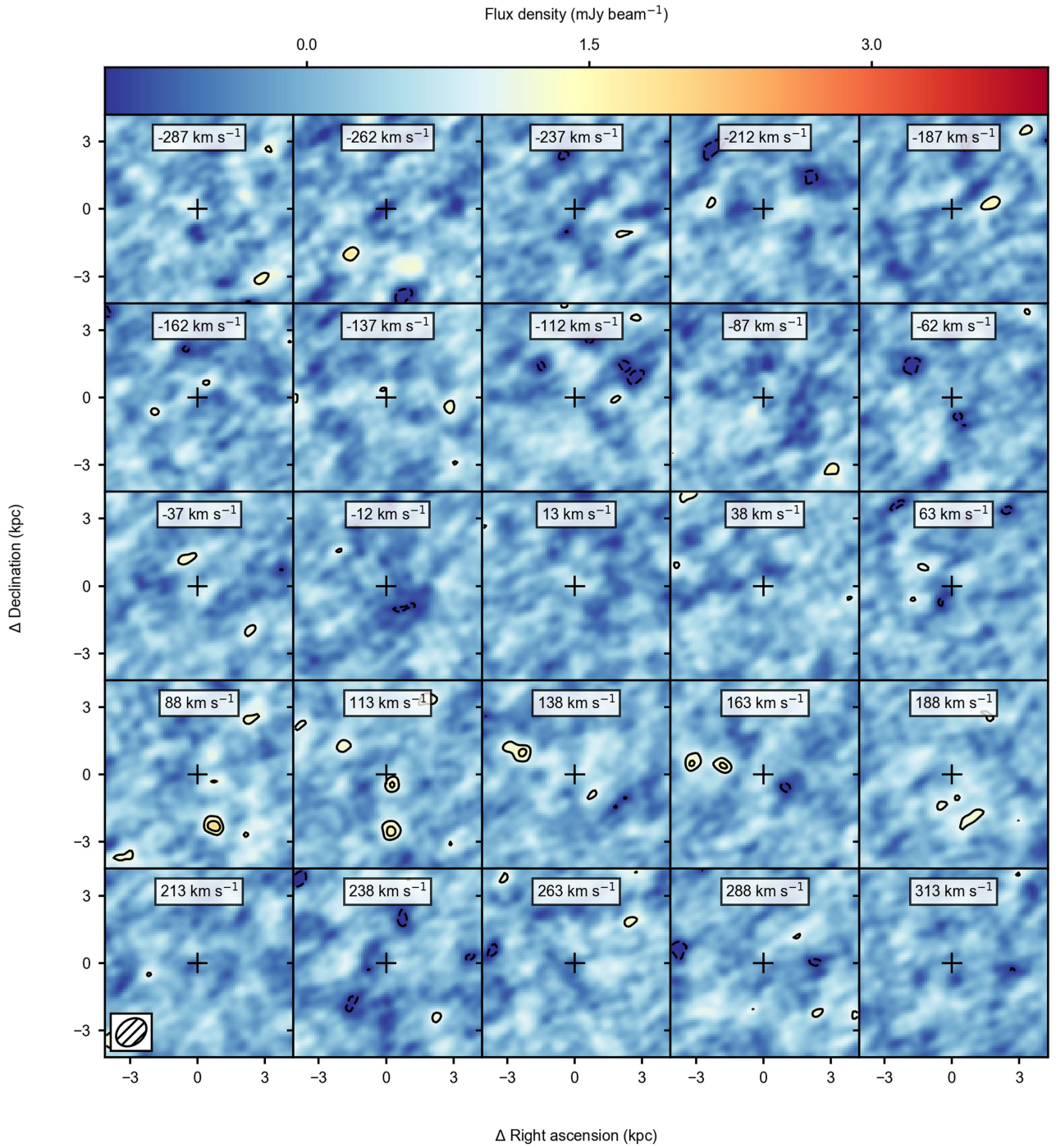


Extended Data Fig. 1 | Spectrum of the [C II] line from DLA0817g. The velocity is relative to the systemic velocity of the [C II] line. The velocity range used to estimate the velocity-integrated [C II] flux density and the integrated [C II] contours is marked by the solid black bar. The 1σ (standard deviation) uncertainty of the measurements is indicated by the dotted red lines, and has been estimated by bootstrapping flux density measurements at random positions within each channel chosen to be devoid of any line emission. A double Gaussian model fit to the data is shown in green. Both the peak flux density of 16.8 ± 1.3 mJy and the velocity-integrated [C II] line flux density of 5.8 ± 0.4 Jy km s⁻¹ are consistent with values obtained from the lower-resolution data, indicating that no emission is resolved out by the higher-resolution observations.



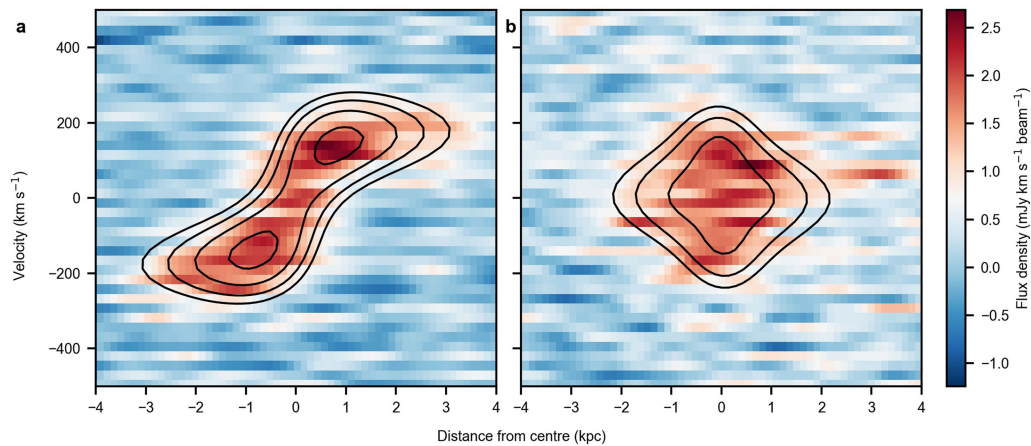
Extended Data Fig. 2 | Channel maps of the [C II] emission line from DLA0817g. The plus symbol indicates the central position of the [C II] emission derived from the kinematic analysis. This agrees within the uncertainties with the position derived from fitting a 2D-Gaussian profile to both the velocity-integrated [C II] emission and the far-infrared continuum emission,

using the *imfit* routine in CASA (Extended Data Table 1). The outer black contour is 3σ , where $\sigma = 0.35 \text{ mJy per beam}$, with subsequent contours increasing in powers of $\sqrt{2}$. Velocities are relative to the kinematically derived [C II] redshift, $z = 4.2603$. The synthesized beam is shown in the bottom left corner of the bottom left panel.



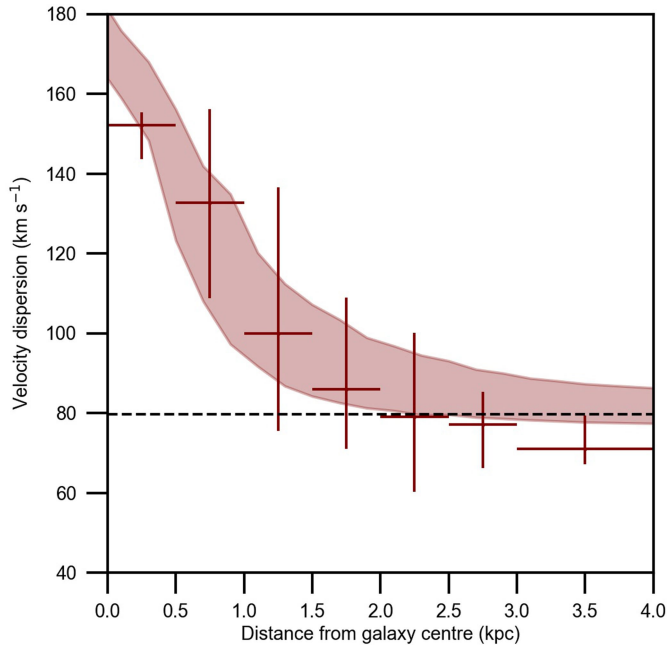
Extended Data Fig. 3 | Channel maps of the residuals, after subtracting the model from the data, of the [C II] emission from DLA0817g. The colour scaling, contour levels and annotations are the same as Extended Data Fig. 2. Little excess emission (at $>3\sigma$ significance) is seen in the individual 25 km s^{-1} channels, indicating that the exponential thin disk model is a good approximation for the bulk of the [C II] emission. Only two features are seen in

the channel maps with $>3\sigma$ emission in two or more consecutive channels: 2.8 kpc south of the centre at 88 km s^{-1} and 113 km s^{-1} , and 3 kpc east of the centre at 113 km s^{-1} , 138 km s^{-1} and 163 km s^{-1} . This emission arises from clumps that are not rotating with the bulk of the gas, possibly arising in outflows or satellite galaxies.

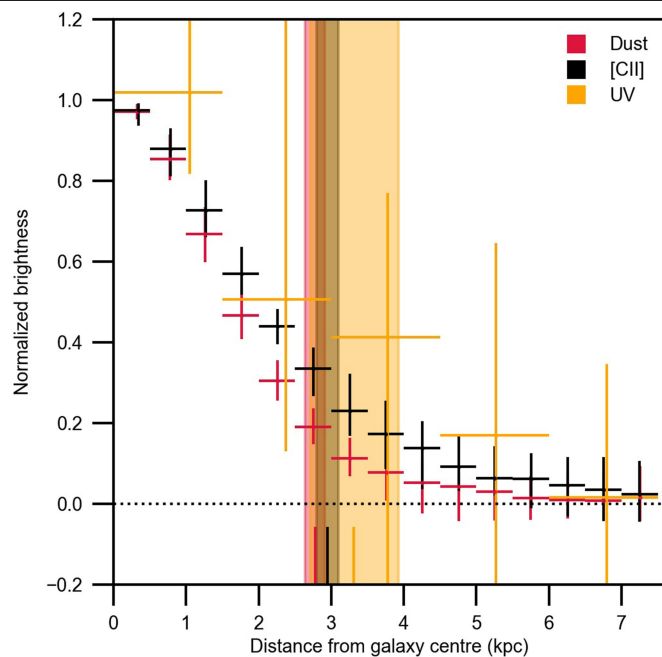


Extended Data Fig. 4 | Position-velocity diagram for DLA0817g. a, The p - v diagram along the major axis of DLA0817g. **b,** The p - v diagram along the minor axis. The contours in both panels are the p - v diagrams derived from the

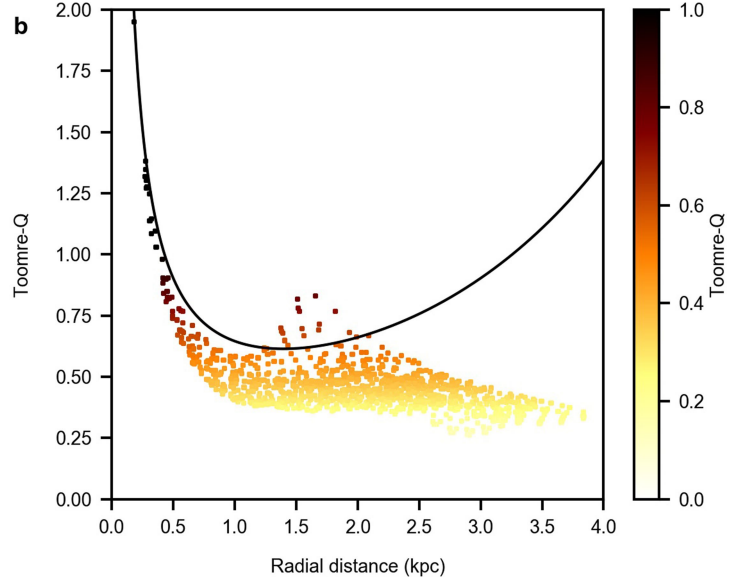
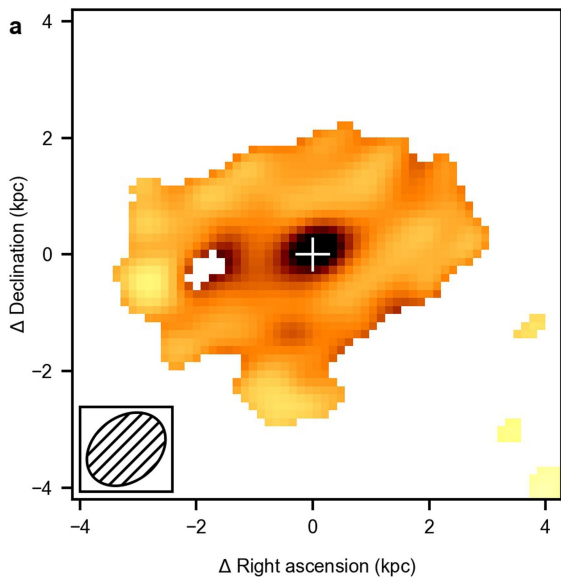
rotating disk model with constant velocity. The outer contour is 2σ where $\sigma = 0.35$ mJy per beam, and the contours increase in powers of $\sqrt{2}$. Distances are given with respect to the kinematic centre of the emission (Fig. 1).



Extended Data Fig. 5 | Velocity dispersion profile for DLA0817g. The observed velocity dispersion profile is measured from the standard deviation of a Gaussian fit to each pixel, and is shown by the data points where the data have been binned into bins equal to the size of the horizontal error bars. The vertical error bars reflect the 16 to 84 percentile spread in measurements per bin. The radius has been de-projected for the inclination of DLA0817g. The dashed line is the value derived from the kinematic modelling. The solid coloured region is the 16 to 84 percentile spread in the constant velocity dispersion model, showing that the increase in velocity dispersion at the galactic centre is due to beam-smearing.



Extended Data Fig. 6 | Light profile for dust continuum, [C II] line and UV emission. The light profiles are scaled by the emission at the kinematic centre. Distances from the kinematic centre are de-projected, taking into account the inclination of the disk. Both the dust continuum and the [C II] emission are convolved with a Gaussian kernel to the slightly worse resolution of the UV observations. This increases the width of the surface density profile by about 10%. Vertical error bars give the 16 to 84 percentile range in the measurements within the bins defined by the horizontal error bars. The [C II] surface density profile, the dust continuum profile and the UV surface density profile—within the uncertainties—are consistent with each other. This can also be seen by the effective radii (marked by solid vertical lines), whose 1σ uncertainties (marked by the vertical coloured regions) overlap at a value of approximately 3 kpc.



Extended Data Fig. 7 | The Toomre- Q parameter for DLA0817g. **a**, The spatial distribution of the Toomre- Q parameter, assuming that the gas density is traced by the [C II] emission. This spatial distribution is still convolved with the ALMA synthesized beam. Over the entire disk, Q is roughly constant and below 1, indicating that the disk is unstable against axisymmetric perturbations. The white cross shows the kinematic centre of the emission, and the inset shows the

ALMA beam for the [C II] emission. **b**, The radial profile of the Toomre- Q parameter. The solid dark line shows Q , assuming that the gas density falls off exponentially, at the same rate as the [C II] emission. The coloured squares are the observed data, as in **a**, corrected for the projected radius. This panel shows that the observed data underestimate Q at large radii owing to beam smearing, which increases the emission, and thus the gas surface density.

Extended Data Table 1 | Physical properties of DLA0817g

[C II] right ascension, R. A. _[C II]	(J2000)	08:17:40.8676(18)
[C II] declination, Decl. _[C II]	(J2000)	+13:51:38.219(19)
[C II] flux density size, $A_{[C II]}$	("×")	$(0.63 \pm 0.07 \times 0.43 \pm 0.05)$
[C II] position angle, P. A. _[C II]	(°)	109 ± 13
[C II] redshift, $z_{[C II]}$		4.2603 ± 0.004
Peak [C II] flux, $S_{[C II], \text{peak}}$	(mJy)	16.8 ± 1.3
[C II] line width, FWHM _[C II]	(km s ⁻¹)	400 ± 40
Integrated [C II] flux, $\int S_{[C II]} dv$	(Jy km s ⁻¹)	(5.8 ± 0.4)
[C II] luminosity, $L_{[C II]}$	(L_{\odot})	$(3.26 \pm 0.22) \times 10^9$
Continuum right ascension, R.A. _{cont}	(J2000)	08:17:40.8655(12)
Continuum declination, Decl. _{cont}	(J2000)	+13:51:38.223(10)
Continuum size, A_{cont}	("×")	$(0.41 \pm 0.05 \times 0.24 \pm 0.03)$
Continuum position angle, P.A. _{cont}	(°)	98 ± 11
Continuum flux, S_{cont}	(mJy)	1.28 ± 0.15
Total infrared luminosity, L_{TIR}	(L_{\odot})	$1.2^{+2.3}_{-0.7} \times 10^{12}$
1.6 μm right ascension, R.A. _{1.6μm}	(J2000)	08:17:40.852(12)
1.6 μm declination, Decl. _{1.6μm}	(J2000)	+13:51:38.19(9)
1.6 μm size, $A_{1.6μm}$	("×")	$(1.2 \pm 0.5 \times 0.26 \pm 0.14)$
1.6 μm position angle, P.A. _{1.6μm}	(°)	114 ± 8
1.6 μm AB magnitude, M_{AB}		25.1 ± 0.2
CO(2–1) right ascension, R.A. _{CO}	(J2000)	08:17:40.53
CO(2–1) declination, Decl. _{CO}	(J2000)	+13:51:34.6
CO(2–1) redshift, z_{CO}		4.2589 ± 0.0010
Integrated CO(2–1) flux, $\int S_{\text{CO}} dv$	(Jy km s ⁻¹)	0.14 ± 0.04
CO(2–1) luminosity, L_{CO}	(L_{\odot})	$(9.6 \pm 2.8) \times 10^6$
CO(2–1) luminosity, L'_{CO}	(K km s ⁻¹ pc ²)	$(2.4 \pm 0.7) \times 10^{10}$
Dust-obscured SFR, $\text{SFR}_{160\mu\text{m}}$	($M_{\odot} \text{ yr}^{-1}$)	118 ± 14
Unobscured SFR, SFR_{NUV}	($M_{\odot} \text{ yr}^{-1}$)	16 ± 3
Molecular gas mass, M_{mol}	(M_{\odot})	$(8.8 \pm 2.6) \times 10^{10}$
Exponential Scale length, R_{d}	(kpc)	$1.39^{+0.06}_{-0.05}$
Dynamical mass within $3R_{\text{d}}$, M_{dyn}	(M_{\odot})	$(7.2 \pm 2.3) \times 10^{10}$
Average Toomre- Q parameter, \bar{Q}		0.96 ± 0.30
v_{rot}/σ_v		$3.4^{+1.1}_{-0.3}$

Physical positions and sizes are estimated from 2D Gaussian fits to the velocity-integrated [C II] flux density and the far-infrared continuum emission, after deconvolving the ALMA synthesized beam. The redshift of the [C II] emission and the full-width at half-maximum, FWHM_[C II], are derived from a double Gaussian fit to the integrated spectrum (Extended Data Fig. 1). The quantities derived from the CO emission have been corrected for the effects of the cosmic microwave background.

Extended Data Table 2 | Results from the kinematic analysis of DLA0817g

		TDconstantV	TDarctanV
R. A.	(J2000)	08:17:40.8667(4)	08:17:40.8666(4)
Decl.	(J2000)	+13:51:38.230(5)	+13:51:38.228(5)
z		4.26033(6)	4.26033(8)
α	($^{\circ}$)	$105.2^{+1.7}_{-1.7}$	$105.7^{+1.7}_{-2.0}$
i	($^{\circ}$)	42^{+3}_{-8}	42^{+4}_{-8}
I_0	(mJy kpc $^{-2}$)	$3.42^{+0.22}_{-0.22}$	$3.3^{+0.3}_{-0.7}$
R_d	(kpc)	$1.39^{+0.06}_{-0.05}$	$1.43^{+0.13}_{-0.08}$
v_{rot}	(km s $^{-1}$)	272^{+52}_{-13}	295^{+54}_{-18}
σ_v	(km s $^{-1}$)	80^{+4}_{-11}	80^{+13}_{-11}
R_v	(kpc)	—	$0.12^{+0.08}_{-0.05}$

Two models are given, both assuming that the [C II] emission arises from a thin disk, but one assuming a constant velocity profile (TDconstantV) and the other an arctangent velocity profile (TDarctanV). The central position of the kinematic centre together with the position angle of the major axis, α , and the inclination, i , determine the position and orientation of the disk. The intensity of the disk is determined by the central intensity, I_0 , and the exponential scale radius, R_d . Finally, the kinematic information of the gas is contained within the maximum rotation velocity, v_{rot} , and the dispersion of the gas σ_v . The arctangent velocity profile requires an additional scale radius, R_v . The number in parenthesis for the right ascension (R.A.), declination (Decl.) and redshift (z) represents the 1 standard deviation uncertainty in the last digit. For the remaining parameters, the uncertainties are asymmetric, and hence the uncertainties listed are the 16 to 84 percentile ranges of the probability distribution function of each parameter.

Quantum entanglement between an atom and a molecule

<https://doi.org/10.1038/s41586-020-2257-1>

Yiheng Lin^{1,2,3,4}✉, David R. Leibbrandt^{2,5}, Dietrich Leibfried^{2,5} & Chin-wen Chou²

Received: 12 December 2019

Accepted: 2 March 2020

Published online: 20 May 2020

 Check for updates

Conventional information processors convert information between different physical carriers for processing, storage and transmission. It seems plausible that quantum information will also be held by different physical carriers in applications such as tests of fundamental physics, quantum enhanced sensors and quantum information processing. Quantum controlled molecules, in particular, could transduce quantum information across a wide range of quantum bit (qubit) frequencies—from a few kilohertz for transitions within the same rotational manifold¹, a few gigahertz for hyperfine transitions, a few terahertz for rotational transitions, to hundreds of terahertz for fundamental and overtone vibrational and electronic transitions—possibly all within the same molecule. Here we demonstrate entanglement between the rotational states of a $^{40}\text{CaH}^+$ molecular ion and the internal states of a $^{40}\text{Ca}^+$ atomic ion². We extend methods used in quantum logic spectroscopy^{1,3} for pure-state initialization, laser manipulation and state readout of the molecular ion. The quantum coherence of the Coulomb coupled motion between the atomic and molecular ions enables subsequent entangling manipulations. The qubit addressed in the molecule has a frequency of either 13.4 kilohertz¹ or 855 gigahertz³, highlighting the versatility of molecular qubits. Our work demonstrates how molecules can transduce quantum information between qubits with different frequencies to enable hybrid quantum systems. We anticipate that our method of quantum control and measurement of molecules will find applications in quantum information science, quantum sensors, fundamental and applied physics, and controlled quantum chemistry.

Quantum state control of atoms has enabled high-fidelity entangling gates for large-scale quantum computation^{4–6}, quantum simulations^{7,8} and multipartite entanglement generation^{9–12}. By adding vibrational and rotational degrees of freedom, as well as coupling of multiple angular momenta to the internal state structure, molecules offer unique opportunities in quantum information processing¹³, precision measurements^{14–17} and tests of fundamental physics¹⁸. Following ideas inspired by laser cooling, trapping and quantum state control of atoms, quantum control of molecules has made substantial progress. Cold atoms have been associated to produce cold molecular ensembles and single molecules^{19,20}. Sub-Doppler laser cooling of the translational motion²¹ and initialization of rotational and vibrational states of molecules^{22–24} have been demonstrated, and dipolar and chemical interactions between molecules^{25–27}, as well as resonant atom–molecule cold collisions²⁸, have been explored. For trapped molecular ions, precision spectroscopy of rotational and vibrational energy levels has been demonstrated^{29,30} and quantum logic spectroscopy (QLS)³¹ has been introduced as an alternative to techniques pioneered on neutral atoms for state detection^{32,33}, preparation and manipulation of molecular ions¹.

To fully realize the potential applications of molecules in quantum science, demonstration of entanglement involving an individually

controlled molecule is a necessary and critical step. Recent proposals suggest using neutral molecules and molecular ions to form qubits^{13,34,35}, and explore their electric dipole moments for long-range interactions. Molecules can also facilitate the construction of a hybrid quantum system. For example, molecules with permanent electric dipole moments can serve as antennas for coupling to quantum systems of disparate nature at very different frequencies, including mechanical cantilevers³⁶ and microwave photons in a superconducting cavity³⁷. For coupled atomic–molecular ion systems³⁸, an atomic ion can also function as a means to prepare entangled states of several co-trapped molecular ions that can be used for quantum enhanced metrology and sensing over the wide frequency range covered by molecules. If control based on QLS can be extended to vibrational transitions and their overtones, a molecular ion can serve as a bridge to connect atomic ion qubits to many other systems, for example, electromagnetic radiation with frequency up to several hundred terahertz, which includes low-transmission-loss photonic (flying) qubits in the 1.5 μm –1.6 μm telecom wavelength range.

Here we leverage QLS to entangle different rotational levels of a molecular ion with an atomic ion, extending previous work^{1,3} on molecular-state preparation, detection and single-qubit control. We trap a single molecular ion alongside a well controlled atomic ion.

¹CAS Key Laboratory of Microscale Magnetic Resonance, Department of Modern Physics, University of Science and Technology of China, Hefei, China. ²Time and Frequency Division, National Institute of Standards and Technology, Boulder, CO, USA. ³Hefei National Laboratory for Physical Sciences at the Microscale, University of Science and Technology of China, Hefei, China.

⁴Synergetic Innovation Center of Quantum Information and Quantum Physics, University of Science and Technology of China, Hefei, China. ⁵Department of Physics, University of Colorado, Boulder, CO, USA. ✉e-mail: yiheng@ustc.edu.cn

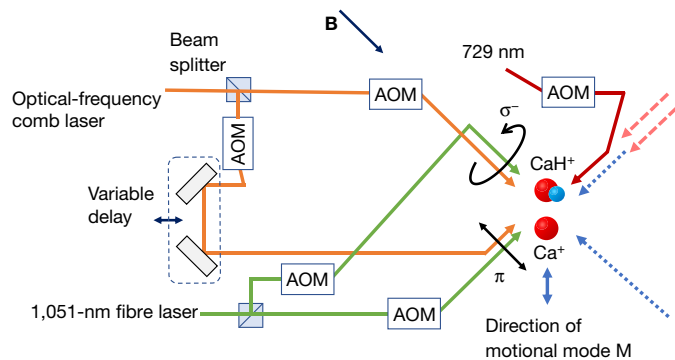


Fig. 1 | Schematic of the experiment. Acousto-optic modulators (AOMs) are used to control the intensities, frequencies and relative phases of the light fields applied on the ions. Transitions between the $|S\rangle$ and $|D\rangle$ states of the $^{40}\text{Ca}^+$ ion are driven by a laser at 729 nm (red solid line). Other laser beams, displayed on the right with dashed arrows, are used for cooling, manipulation and state detection of the $^{40}\text{Ca}^+$ ion (see Methods). The molecule is manipulated by driving two-photon stimulated Raman transitions with two beams generated from a continuous-wave 1,051-nm fibre laser (green lines) and two beams generated by an optical-frequency-comb laser centred at 850 nm (orange lines). The pairs of beams are offset in frequency by two AOMs and polarized circularly (σ^-) and linearly (π). The σ^- beams are along the direction of the magnetic field (\mathbf{B}), which is $\sim 45^\circ$ from the line connecting the equilibrium locations of the ions. The variable delay is adjusted to ensure that the ~ 40 -fs pulses of the two beams from the frequency comb overlap temporally on the molecule. The direction of the motional mode M is along the line connecting the ion equilibrium positions.

The atom is used to laser-cool the coupled translational harmonic motion of the two ions, prepare a pure quantum state of the molecule¹ and serve as a high-fidelity qubit² in our entanglement demonstration. We then apply tailored laser pulse sequences to generate entangled states of the rotation of the molecular ion and the long-lived 411.0-THz (729 nm) electronic qubit of the atomic ion. We achieve an entangled-state fidelity of 0.87(3) for a molecular qubit within a rotational manifold with a frequency of ~ 13.4 kHz, and a fidelity of 0.76(3) for a qubit in different rotational manifolds with a much higher frequency of approximately 855 GHz (all uncertainties reported here are one standard deviation). In both cases, we obtain fidelities much greater than the threshold for genuine two-partite entanglement of $1/2$ (ref. 39). Our demonstration provides evidence that atoms can be entangled with various molecular rotational states, with their frequency differences spanning more than seven orders of magnitude.

In our experiments, a $^{40}\text{Ca}^+$ atomic ion is co-trapped with a $^{40}\text{CaH}^+$ molecular ion in a linear Paul trap¹ (see Fig. 1). A static external magnetic field \mathbf{B} with magnitude ~ 0.36 mT provides a quantization axis. The Coulomb repulsion between the ions results in two normal modes of coupled harmonic motion along each of three orthogonal directions, with the two ions moving in phase or out of phase. These modes of coupled motion are cooled by applying lasers that are all near-resonant with transitions in the $^{40}\text{Ca}^+$ ion⁴⁰ (see Methods). To transfer and manipulate quantum states of the co-trapped molecular ion by QLS³¹, we utilize the out-of-phase mode M at ~ 5.16 MHz along the direction connecting the equilibrium positions of the ions. The quantized state with n phonons (motional quanta) of this mode is denoted as $|n\rangle_M$. For ground-state cooling of the coupled motional modes and preparation of entangled states, in the $^{40}\text{Ca}^+$ atom we use the ground electronic state $|S\rangle \equiv |S_{1/2}, m_j = -1/2\rangle$ and a metastable excited state $|D\rangle \equiv |D_{5/2}, m_j = -5/2\rangle$ with a lifetime of approximately 1 s (see Fig. 2a), where m_j is the quantum number of the component of the electronic angular momentum along \mathbf{B} . These atomic qubit states are coupled by driving an electric quadrupole transition with a laser near 729 nm (see Fig. 1) and can be distinguished by state-dependent fluorescence detection (see Methods).

The molecule is in its Σ electronic and vibrational ground state at room temperature. The rotational eigenstates in the presence of \mathbf{B} are denoted as $|J, m, \xi\rangle$, where the non-negative integer J is the rotational angular momentum quantum number; $m = m_j + m_l$ is the sum of the components of the rotational angular momentum ($-J \leq m_j \leq J$) and the proton spin ($m_l = \pm 1/2$) along \mathbf{B} ; and $\xi = \{+, -\}$ distinguishes the two eigenstates with the same m that are split in energy owing to the interaction between the rotational angular momentum and the nuclear spin, except for the stretch states with $m = \pm(J + 1/2)$, where ξ indicates the sign of m (ref. 1). States within a rotational manifold with the same J are manipulated by driving stimulated Raman transitions with light fields from a fibre laser at 1,051 nm (ref. 1), as schematically shown in Fig. 1. In particular, we can tune the frequency difference between these light fields to address and initialize a low-frequency molecular qubit composed of two states within the $J = 2$ manifold, $|2, -3/2, -\rangle \equiv |-3/2\rangle$ and $|2, -5/2, -\rangle \equiv |-5/2\rangle$, with transition frequency ~ 13.4 kHz (see Fig. 2b).

We start the entanglement sequence by preparing the $^{40}\text{CaH}^+$ ion in the $|-3/2\rangle$ state in a probabilistic but heralded fashion^{1,3} (also see Methods). Subsequently, we apply ground-state cooling and optical pumping on the $^{40}\text{Ca}^+$ ion, ideally leaving the system in the state

$$|\Psi_0\rangle = |S\rangle |-3/2\rangle |0\rangle_M \quad (1)$$

The target entangled state of the low-frequency molecular qubit with the atom has the form

$$|\psi\rangle = \frac{1}{\sqrt{2}}(|S\rangle |-3/2\rangle + |D\rangle |-5/2\rangle) \quad (2)$$

This state consists of a superposition in which the lower energy state of the atom and the higher energy state of the molecule are paired and vice versa. The entangled state $|\psi\rangle$ therefore has odd parity. Starting with $|\Psi_0\rangle$, we drive a $\pi/2$ pulse on the molecular Raman sideband transition $|-3/2\rangle |0\rangle_M \leftrightarrow |-5/2\rangle |1\rangle_M$ (see Methods) to ideally prepare

$$|\Psi_i\rangle = \frac{1}{\sqrt{2}}(|S\rangle |-3/2\rangle |0\rangle_M + |D\rangle |-5/2\rangle |1\rangle_M) \quad (3)$$

The intermediate state $|\Psi_i\rangle$ is an entangled state between the molecular qubit and the mode of motion M . We transfer this entanglement from the motion to the $^{40}\text{Ca}^+$ atom by driving a π pulse on its $|S\rangle |1\rangle_M \leftrightarrow |D\rangle |0\rangle_M$ sideband transition. This pulse of the 729-nm laser does not affect the $|S\rangle |-3/2\rangle |0\rangle_M$ component of $|\Psi_i\rangle$, whereas the $|S\rangle |-5/2\rangle |1\rangle_M$ component ends in the state $|D\rangle |-5/2\rangle |0\rangle_M$. In this way, the final motional state $|0\rangle_M$ is a common factor of both components of the superposition that multiplies the desired entangled state $|\psi\rangle$ of the atom and the molecule. We start the pulse sequence on the molecule, which has zero electron spin and therefore a weak dependence of its energy levels on the external magnetic field, to reduce effects of the relatively short (~ 1 ms) coherence time of the $^{40}\text{Ca}^+$ qubit, which is limited by electron-spin couplings to magnetic-field fluctuations in our setup.

We characterize the entangled state with measurements of the state populations P_{ζ} (the probability of finding the atom and the molecule in the state $|\zeta\rangle$) within the four-state subspace $\{|S\rangle |-5/2\rangle, |D\rangle |-3/2\rangle, |S\rangle |-3/2\rangle, |D\rangle |-5/2\rangle\}$ of the atom and the molecule, and by characterizing the coherence between the states⁴¹. We determine P_{ζ} by applying state-dependent fluorescence detection on the atomic states, and subsequently detecting the molecular states by transferring them to the atom via the motional mode M with quantum logic, all in the same experiment trial (see Methods). Repeating the sequence of entanglement generation, followed by atomic- and molecular-state detection, accumulates statistics for P_{ζ} . The coherence of the entangled state can be characterized by applying an additional ‘analysis’ $\pi/2$ pulse to the atomic qubit and a $\pi/2$ pulse to the molecular qubit after the state is created, with a variable phase of ϕ_a and $-\phi_a$.

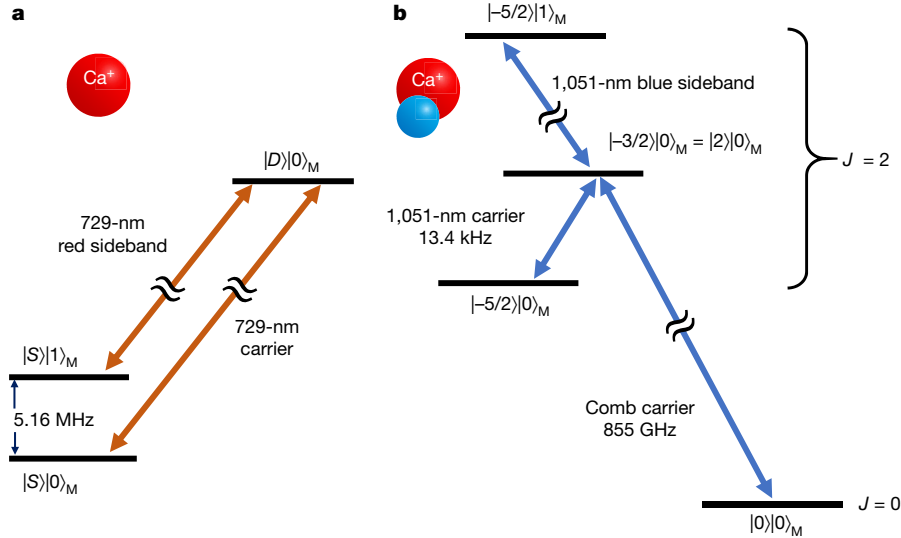


Fig. 2 | Energy levels and selected laser-driven transitions. a, b, Energy levels and transitions are shown for the $^{40}\text{Ca}^+$ atomic ion (a) and the $^{40}\text{CaH}^+$ molecular ion (b). Here ‘carrier’ denotes transitions between ion states that do not change the state $|n\rangle_M$ of the out-of-phase motional mode M, whereas ‘sideband’

transitions add or subtract one motional quantum for this mode and change the ion state. As described in more detail in the text, $|S\rangle$ and $|D\rangle$ are electronic states of the atom, and $|-5/2\rangle, |-3/2\rangle = |2\rangle$ and $|0\rangle$ denote rotational states of the molecule.

respectively, relative to the pulses used during state creation⁴¹. This leads to interference between the superposition parts of the entangled state, which is reflected in the populations $P_i(\phi_a)$ observed after the $\pi/2$ pulses. In particular, the parity

$$\Pi(\phi_a) = P_{S,-5/2}(\phi_a) + P_{D,-3/2}(\phi_a) - [P_{S,-3/2}(\phi_a) + P_{D,-5/2}(\phi_a)] \quad (4)$$

oscillates as $C\cos(2\phi_a + \phi_0)$, where ϕ_0 is an offset in phase and $C \geq 0$ is the observed contrast. The fidelity between the entangled state produced in the experiment and $|\psi\rangle$ is then⁴¹

$$F_1 = \frac{1}{2}(P_{S,-3/2} + P_{D,-5/2} + C)$$

Figure 3a shows the observed parity signal (circles) plotted versus the analysis phase ϕ_a , with on average -99 realizations of the entangled state at each phase (see Methods). The solid line is a cosine fit of $C\cos(2\phi_a + \phi_0)$ to the observed parity fringe, with contrast $C = 0.78(4)$. Along with the populations $P_{S,-3/2} = 0.50(4)$ and $P_{D,-5/2} = 0.45(4)$ obtained directly from 202 experimental realizations for state $|\psi\rangle$ (without applying the analysis pulses), this yields a fidelity of $F_1 = 0.87(3) > 0.5$ indicating bipartite entanglement³⁹. This is a lower bound for the actual fidelity of the prepared state because it also includes the infidelity introduced by imperfect readout. Additional known reductions in entangled-state fidelity arise from imperfections in ground-state cooling, non-ideal initial molecular-state preparation into $|-3/2\rangle$, residual nonlinear coupling between different normal modes of motion, and off-resonant coupling to nearby transitions in the molecule.

To demonstrate the versatility of molecules, we also entangle the atom with a high-frequency molecular qubit composed of $|2\rangle \equiv |2, -3/2, -\rangle = |-3/2\rangle$ (a state shared with the low-frequency qubit) and $|0\rangle \equiv |0, -1/2, -\rangle$ with a transition frequency of -855 GHz (see Fig. 2b and Methods). The target atom–molecule entangled state in this case has the form

$$|\psi_h\rangle = \frac{1}{\sqrt{2}}(|D\rangle|2\rangle + |S\rangle|0\rangle) \quad (5)$$

The lower energy states and the higher energy states of the two ions are paired, making $|\psi_h\rangle$ an even-parity state. We manipulate the high-frequency molecular qubit using stimulated Raman transitions

induced by an optical-frequency comb, as described theoretically in refs.^{42,43} and demonstrated experimentally in ref.³. The two beams originate from the same source, with the frequency of each beam shifted by an acousto-optic modulator to match the frequency differences of pairs of comb teeth with the transition frequency of the molecule to collectively drive the corresponding Raman transition (see Fig. 1 and Methods). After initial preparation of the system in the intermediate state $|\psi_i\rangle$, we map $|-3/2\rangle|0\rangle_M = |2\rangle|0\rangle_M$ to $|0\rangle|0\rangle_M$ with a carrier π pulse of the comb laser, followed by a carrier π pulse from the 1,051-nm laser that maps $|-5/2\rangle|1\rangle_M$ to $|2\rangle|1\rangle_M$. A subsequent $|S\rangle|1\rangle_M \leftrightarrow |D\rangle|0\rangle_M$ sideband π pulse on the atom ideally prepares $|\psi_h\rangle$.

To quantify the fidelity F_h between $|\psi_h\rangle$ and the experimentally realized entangled state with the high-frequency molecular qubit, population measurements are conducted in a similar way as for the entangled state involving the low-frequency molecular qubit. To find the contrast of the parity fringe, we need to apply a $\pi/2$ pulse with the frequency comb to address the high-frequency molecular qubit (see Methods). By scanning the analysis phase ϕ_a for the $\pi/2$ pulse of the comb and for the 729-nm $\pi/2$ pulse applied to the atom in equal steps with the same sign, we obtain the signal shown in Fig. 3b with

$$\Pi_h(\phi_a) = P_{S,0}(\phi_a) + P_{D,2}(\phi_a) - [P_{S,2}(\phi_a) + P_{D,0}(\phi_a)] \quad (6)$$

The signs with which the phases of the analysis $\pi/2$ pulses are scanned for the different entangled states arise from the opposite parity for the entangled components in the states $|\psi\rangle$ and $|\psi_h\rangle$ (see equations (2) and (5)), which highlights the nature of mixed-species entanglement and the versatility in molecular qubit states. A fit to the parity signal (fitted contrast of 0.65(5), with on average -79 realizations of $|\psi_h\rangle$ per phase angle ϕ_a) together with population measurements ($P_{S,0} = 0.47(2)$ and $P_{D,2} = 0.40(2)$, averaged over 491 realizations of $|\psi_h\rangle$) yields $F_h = 0.76(3)$. We attribute the decrease in fidelity with respect to that of $|\psi\rangle$ mainly to the finite coherence of the frequency comb and the larger number of imperfect operations in the pulse sequence used to produce $|\psi_h\rangle$.

Ways to improve the entanglement fidelity include better ground-state cooling and reducing the nonlinear cross-coupling between motional modes^{44,45}. Qubit decoherence can cause deviation of the experimentally realized state from the target pure state

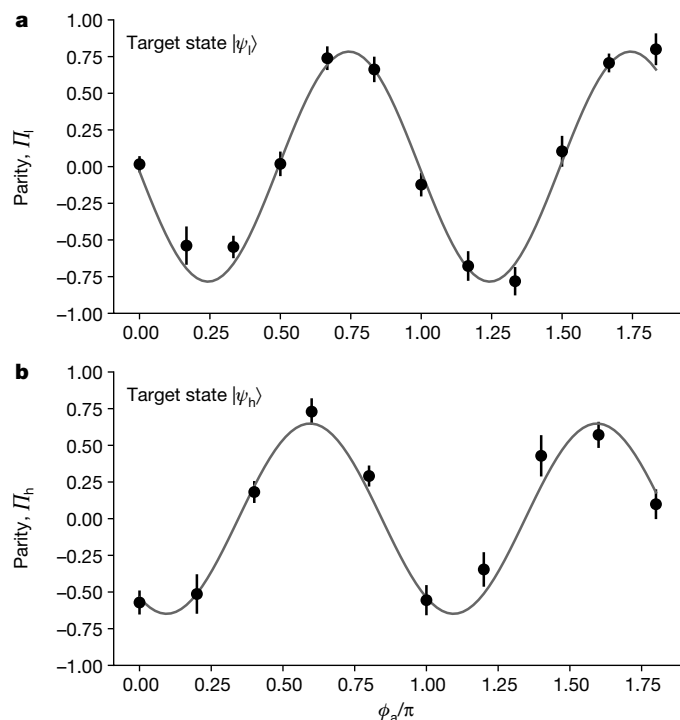


Fig. 3 | Parity measurements of the entangled states. **a**, Parity fringe (see equation (4)) deduced from the measured populations after applying $\pi/2$ pulses to the experimentally prepared entangled state $|\psi_i\rangle$ of the atom and the low-frequency molecular qubit with frequency -13.4 kHz. The phase ϕ_a is scanned using equal and opposite steps on the atomic and the molecular ions, respectively. We observe a sinusoidal parity fringe contrast of $C = 0.78(4)$ from a least-squares fit to the data, weighted by their statistical standard deviation of the mean. **b**, Parity fringe (see equation (6)) deduced from the populations measured after applying $\pi/2$ pulses to the experimentally realized entangled state $|\psi_h\rangle$ of the atom and the high-frequency molecular qubit with frequency -855 GHz. Here ϕ_a is scanned with equal steps on the two ions. We observe a parity fringe contrast of $C = 0.65(5)$. The offsets of the initial phase ϕ_0 in both graphs are caused by residual Stark shifts induced by the 729-nm beam and the 1,051-nm beams (for **a**) and the comb beams (for **b**). Error bars denote one standard deviation from the mean.

and decoherence of the entangled states. The main cause of dephasing of the entangled states in our experiments is the variation of the $^{40}\text{Ca}^+$ qubit frequency due to magnetic-field fluctuations. For the high-frequency molecular qubit, the few-millisecond coherence time of the frequency comb leads to a substantial loss in contrast of the parity curve and could be improved by several orders of magnitude with better control of the repetition rate⁴⁶. A cryogenic ion trap would suppress blackbody radiation and further increase the lifetime and coherence time of molecular rotational states. At the same time, the vacuum would be greatly improved, reducing the rate of collisions of the ions with background gas that can lead to perturbation of the states and to ions trading places. Atomic qubit coherence can be improved by better stabilization of the magnetic field⁴⁷ and by choosing a qubit that is less sensitive to magnetic-field fluctuations⁴⁸. With such improvements, higher fidelity would ensue and the coherence time of the entangled states could also be lengthened.

In summary, we use elements of quantum logic spectroscopy and coherent manipulation of the resulting pure quantum states to create and characterize entanglement between long-lived electronic states of a $^{40}\text{Ca}^+$ atom and rotational states of a $^{40}\text{CaH}^+$ molecular ion, trapped together in the same potential well. We demonstrate entanglement between an atomic qubit of frequency 411.0 THz (729 nm) and molecular rotational qubits, connected by transitions of frequencies at either

13.4 kHz or 855 GHz, establishing the suitability of molecules for quantum state transduction between qubits of very different frequencies. We observe fidelities of the entangled states of 0.87(3) and 0.76(3), respectively. Our experimental approach is suitable for a wide range of molecular ions, offering a broad selection of qubit frequencies and properties. In particular, stimulated Raman transitions can also be driven in symmetric diatomic molecules, which have no permanent electric dipole moment and could provide qubit coherence times longer than a few minutes. This work shows that entanglement involving the quantum states of a molecule is feasible and offers versatility that may be useful for a range of applications.

Online content

Any methods, additional references, Nature Research reporting summaries, source data, extended data, supplementary information, acknowledgements, peer review information; details of author contributions and competing interests; and statements of data and code availability are available at <https://doi.org/10.1038/s41586-020-2257-1>.

1. Chou, C.-w. et al. Preparation and coherent manipulation of pure quantum states of a single molecular ion. *Nature* **545**, 203–207 (2017).
2. Bermudez, A. et al. Assessing the progress of trapped-ion processors towards fault-tolerant quantum computation. *Phys. Rev. X* **7**, 041061 (2017).
3. Chou, C.-w. et al. Frequency-comb spectroscopy on pure quantum states of a single molecular ion. *Science* **367**, 1458–1461 (2020).
4. Ballance, C. J., Harty, T. P., Linke, N. M., Sepiol, M. A. & Lucas, D. M. High-fidelity quantum logic gates using trapped-ion hyperfine qubits. *Phys. Rev. Lett.* **117**, 060504 (2016).
5. Gaebler, J. P. et al. High-fidelity universal gate set for $^{9}\text{Be}^+$ ion qubits. *Phys. Rev. Lett.* **117**, 060505 (2016).
6. Levine, H. et al. High-fidelity control and entanglement of Rydberg-atom qubits. *Phys. Rev. Lett.* **121**, 123603 (2018).
7. Bernien, H. et al. Probing many-body dynamics on a 51-atom quantum simulator. *Nature* **551**, 579–584 (2017).
8. Zhang, J. et al. Observation of a many-body dynamical phase transition with a 53-qubit quantum simulator. *Nature* **551**, 601–604 (2017).
9. McConnell, R., Zhang, H., Hu, J., Čuk, S. & Vuletić, V. Entanglement with negative Wigner function of almost 3,000 atoms heralded by one photon. *Nature* **519**, 439–442 (2015).
10. Bohnet, J. G. et al. Quantum spin dynamics and entanglement generation with hundreds of trapped ions. *Science* **352**, 1297–1301 (2016).
11. Luo, X.-Y. et al. Deterministic entanglement generation from driving through quantum phase transitions. *Science* **355**, 620–623 (2017).
12. Omran, A. et al. Generation and manipulation of Schrödinger cat states in Rydberg atom arrays. *Science* **365**, 570–574 (2019).
13. DeMille, D. Quantum computation with trapped polar molecules. *Phys. Rev. Lett.* **88**, 067901 (2002).
14. Kozryyev, I. & Hutzler, N. R. Precision measurement of time-reversal symmetry violation with laser-cooled polyatomic molecules. *Phys. Rev. Lett.* **119**, 133002 (2017).
15. Cairncross, W. B. et al. Precision measurement of the electron's electric dipole moment using trapped molecular ions. *Phys. Rev. Lett.* **119**, 153001 (2017).
16. Altıntaş, E., Ammon, J., Cahn, S. B. & DeMille, D. Demonstration of a sensitive method to measure nuclear-spin-dependent parity violation. *Phys. Rev. Lett.* **120**, 142501 (2018).
17. ACME Collaboration. Improved limit on the electric dipole moment of the electron. *Nature* **562**, 355–360 (2018).
18. Saffronova, M. S. et al. Search for new physics with atoms and molecules. *Rev. Mod. Phys.* **90**, 025008 (2018).
19. Moses, S. A., Covey, J. P., Mieczkowski, M. T., Jin, D. S. & Ye, J. New frontiers for quantum gases of polar molecules. *Nat. Phys.* **13**, 13–20 (2017).
20. Liu, L. R. et al. Building one molecule from a reservoir of two atoms. *Science* **360**, 900–903 (2018).
21. Truppe, S. et al. Molecules cooled below the Doppler limit. *Nat. Phys.* **13**, 1173–1176 (2017).
22. Ospelkaus, S. et al. Controlling the hyperfine state of rovibronic ground-state polar molecules. *Phys. Rev. Lett.* **104**, 030402 (2010).
23. Reinaudi, G., Osborn, C. B., McDonald, M., Kotochigova, S. & Zelevinsky, T. Optical production of stable ultracold $^{88}\text{Sr}_2$ molecules. *Phys. Rev. Lett.* **109**, 115303 (2012).
24. Park, J. W., Will, S. A. & Zwierlein, M. W. Ultracold dipolar gas of fermionic $^{23}\text{Na}^{40}\text{K}$ molecules in their absolute ground state. *Phys. Rev. Lett.* **114**, 205302 (2015).
25. Yan, B. et al. Observation of dipolar spin-exchange interactions with lattice-confined polar molecules. *Nature* **501**, 521–525 (2013).
26. Hu, M.-G. et al. Direct observation of bimolecular reactions of ultracold KRB molecules. *Science* **366**, 1111–1115 (2019).
27. Cheuk, L. W. et al. Observation of collisions between two ultracold ground-state CaF molecules. Preprint at <http://arxiv.org/abs/2002.00048> (2020).
28. Yang, H. et al. Observation of magnetically tunable Feshbach resonances in ultracold $^{23}\text{Na}^{40}\text{K} + ^{40}\text{K}$ collisions. *Science* **363**, 261–264 (2019).
29. Biesheuvel, J. et al. Probing QED and fundamental constants through laser spectroscopy of vibrational transitions in HD^+ . *Nat. Commun.* **7**, 10385 (2016).
30. Alighanbari, S., Hansen, M. G., Korobov, V. I. & Schiller, S. Rotational spectroscopy of cold and trapped molecular ions in the Lamb–Dicke regime. *Nat. Phys.* **14**, 555–559 (2018).

31. Schmidt, P. O. et al. Spectroscopy using quantum logic. *Science* **309**, 749–752 (2005).
32. Wolf, F. et al. Non-destructive state detection for quantum logic spectroscopy of molecular ions. *Nature* **530**, 457–460 (2016).
33. Sinhal, M., Meir, Z., Najafian, K., Hegi, G. & Willitsch, S. Quantum non-demolition state detection and spectroscopy of single trapped molecules. *Science* **367**, 1213–1218 (2020).
34. Hudson, E. R. & Campbell, W. C. Dipolar quantum logic for freely rotating trapped molecular ions. *Phys. Rev. A* **98**, 040302 (2018).
35. Ni, K.-K., Rosenband, T. & Grimes, D. D. Dipolar exchange quantum logic gate with polar molecules. *Chem. Sci.* **9**, 6830–6838 (2018).
36. Kippenberg, T. J. & Vahala, K. J. Cavity optomechanics: back-action at the mesoscale. *Science* **321**, 1172–1176 (2008).
37. Schuster, D. I., Bishop, L. S., Chuang, I. L., DeMille, D. & Schoelkopf, R. J. Cavity QED in a molecular ion trap. *Phys. Rev. A* **83**, 012311 (2011).
38. Campbell, W. C. & Hudson, E. R. Dipole-phonon quantum logic with trapped polar molecular ions. Preprint at <http://arxiv.org/abs/1909.02668> (2019).
39. Leibfried, D. et al. Creation of a six-atom ‘Schrödinger cat’ state. *Nature* **438**, 639–642 (2005).
40. Rugango, R. et al. Sympathetic cooling of molecular ion motion to the ground state. *New J. Phys.* **17**, 035009 (2015).
41. Sackett, C. A. et al. Experimental entanglement of four particles. *Nature* **404**, 256–259 (2000).
42. Ding, S. & Matsukevich, D. N. Quantum logic for the control and manipulation of molecular ions using a frequency comb. *New J. Phys.* **14**, 023028 (2012).
43. Leibfried, D. Quantum state preparation and control of single molecular ions. *New J. Phys.* **14**, 023029 (2012).
44. Roos, C. F. et al. Nonlinear coupling of continuous variables at the single quantum level. *Phys. Rev. A* **77**, 040302 (2008).
45. Nie, X. R., Roos, C. F. & James, D. F. V. Theory of cross phase modulation for the vibrational modes of trapped ions. *Phys. Lett. A* **373**, 422–425 (2009).
46. Bartels, A., Oates, C. W., Hollberg, L. & Diddams, S. A. Stabilization of femtosecond laser frequency combs with subhertz residual linewidths. *Opt. Lett.* **29**, 1081–1083 (2004).
47. Merkel, B. et al. Magnetic field stabilization system for atomic physics experiments. *Rev. Sci. Instrum.* **90**, 044702 (2019).
48. Langer, C. et al. Long-lived qubit memory using atomic ions. *Phys. Rev. Lett.* **95**, 060502 (2005).

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

© The Author(s), under exclusive licence to Springer Nature Limited 2020

Methods

Atomic-state manipulation and detection

We apply laser cooling to $^{40}\text{Ca}^+$ only, using lasers that are near-resonant with atomic transitions to sympathetically cool all motional modes of the two-ion crystal. We perform Doppler cooling of $^{40}\text{Ca}^+$ with laser light at 397 nm driving the transition between the $S_{1/2}$ and $P_{1/2}$ states, and use lasers at 866 nm and 854 nm to repump from the metastable $D_{3/2}$ and $D_{5/2}$ states, respectively. See Fig. 1 for a schematic layout of the beam lines, with blue short-dashed arrows depicting the 397-nm beams and pink long-dashed arrows depicting the 854-nm and 866-nm beams. After Doppler cooling, which cools all motional modes, we prepare the axial modes (modes parallel to the direction connecting the equilibrium positions of the two ions) near their ground state by electromagnetically induced-transparency cooling⁴⁹ and sideband cooling. To drive sideband transitions on the out-of-phase axial mode M, we apply a 729-nm pulse with a duration of 45 μs , well within the qubit coherence time of the atom, during both cooling and creating entangled states. We also apply sideband cooling on the out-of-phase radial modes⁵⁰, along two directions perpendicular to the axial mode and to each other. The out-of-phase radial modes must be cooled close to the ground state to minimize frequency shifts and motional decoherence in the axial out-of-phase mode caused by nonlinear coupling between the radial and axial modes^{44,45}.

During fluorescence detection, we apply a near-resonant 397-nm laser to the $^{40}\text{Ca}^+$ atom driving the transitions between levels in the $S_{1/2}$ and $P_{1/2}$ manifolds together with an 866-nm laser to repump from the $D_{3/2}$ states. The photons scattered by the atom are directed to a photomultiplier tube (PMT) in the setup. On average, over an 85- μs detection window, the PMT registers approximately 20 counts for the $|S\rangle$ state and <0.5 counts for the $|D\rangle$ state.

Molecular-state preparation and detection, Raman transitions and parity measurement of $|\psi_h\rangle$

To prepare the molecule in $|-3/2\rangle$, we start with the $|D\rangle$ state for the atom, cool the motional mode M close to $|0\rangle_M$ and excite it ideally to $|1\rangle_M$ with a π pulse on the atomic sideband implementing $|D\rangle|0\rangle_M \rightarrow |S\rangle|1\rangle_M$. We then apply a π pulse on the molecular sideband that drives $|-5/2\rangle|1\rangle_M \rightarrow |-3/2\rangle|0\rangle_M$, followed by a π pulse on the atomic sideband $|S\rangle|1\rangle_M \rightarrow |D\rangle|0\rangle_M$, which does not affect $|S\rangle|0\rangle_M$. Subsequent atomic-state detection projects the molecular state to $|-3/2\rangle$ with high probability if the detection outcome is $|S\rangle$. This sequence can be repeated while alternating with a sequence preparing $|-5/2\rangle$, until the confidence level for the molecular state being $|-3/2\rangle$ is above a preset threshold.

Detecting whether the molecule is in the $|-3/2\rangle$ state is achieved by preparing the atom and the motional mode M in $|D\rangle|0\rangle_M$, applying a π pulse on the molecular sideband $|-3/2\rangle|0\rangle_M \rightarrow |-5/2\rangle|1\rangle_M$, then a π pulse on the atomic sideband $|D\rangle|1\rangle_M \rightarrow |S\rangle|0\rangle_M$ (which leaves $|D\rangle|0\rangle_M$ unchanged), followed by an atomic-state detection. A detection outcome $|S\rangle$ indicates that the molecule is in the $|-3/2\rangle$ state. The detection outcome $|D\rangle$ is attributed to the molecule being in another molecular qubit state.

To drive Raman sideband transitions and implement $|\psi_0\rangle \rightarrow |\psi_i\rangle$, we apply a 1,051-nm laser pulse with smooth 300- μs rising and falling edges and a plateau with duration of 162.5 μs to avoid sharp pulse edges with increased frequency components driving other molecular transitions off-resonantly.

The spectrum of the optical-frequency comb has a full-width at half-maximum of approximately 20 nm around a centre wavelength of ~850 nm and a repetition rate of $f_{\text{rep}} \approx 80$ MHz. For stimulated Raman transitions driven with optical-frequency-comb beams, we split the comb laser output into two different beams and control their frequency and phase differences using AOMs (Fig. 1), with the same drive frequency f_{AOM} but with opposite diffraction orders. The AOMs allow us

to scan the frequency difference $2f_{\text{AOM}}$ of the beams over a range that exceeds the repetition rate. The comb is far off-resonance from any electronic transition in the molecule, but stimulated Raman transitions at frequency f_{Raman} can be driven simultaneously by all pairs of comb teeth (one comb tooth from each beam for every pair) with matching frequency difference $f_{\text{Raman}} = |Nf_{\text{rep}} - 2f_{\text{AOM}}|$, where N is an integer number with magnitude of order 10,000 (see below) and the sign depends on whether a photon is absorbed from the beam with σ^- or π polarization³.

In the parity analysis for $|\psi_h\rangle$ we need to apply a $\pi/2$ pulse with the frequency comb, which has frequency components close to the $D_{5/2} \leftrightarrow P_{3/2}$ transition of $^{40}\text{Ca}^+$. To avoid affecting the state of the atom with this pulse, we apply sideband π pulses on the atom mapping $|D\rangle|0\rangle_M \rightarrow |S\rangle|1\rangle_M$ to hide the atomic population in $|S\rangle$ before the comb pulse and $|S\rangle|1\rangle_M \rightarrow |D\rangle|0\rangle_M$ afterwards for the population measurements.

Statistics for entangled-state analyses

To determine the fidelity for the realization of $|\psi_i\rangle$, we perform parity measurements with $\phi_a = \frac{\pi}{6} \times \{0, 1, 2, \dots, 11\}$ and number of trials {246, 39, 115, 106, 92, 83, 114, 62, 64, 67, 150, 50}, respectively. To determine the fidelity of the experimentally prepared $|\psi_h\rangle$, we perform parity measurements with $\phi_a = \frac{\pi}{5} \times \{0, 1, 2, \dots, 9\}$ and number of trials {98, 37, 132, 74, 141, 63, 52, 35, 84, 71}, respectively. The variation in the number of trials is the result of repeating the experimental sequence with the same parameters including ϕ_a after we confirm that the molecular state is heralded in the initial state of $|-3/2\rangle$. After each repetition, we check whether the molecule remains in the desired manifold of $\{|-3/2\rangle, |-5/2\rangle\}$, where we design the verification measurement so that a positive result brings the molecule back to the $|-3/2\rangle$ state. If the molecule has left the $\{|-3/2\rangle, |-5/2\rangle\}$ manifold, we randomly draw a new value of ϕ_a from the list for the next time the molecular state is heralded in $|-3/2\rangle$.

Determination of the $^{40}\text{CaH}^+|2\rangle \equiv |2, -3/2, -\rangle \leftrightarrow |0\rangle \equiv |0, -1/2, -\rangle$ transition frequency

The transition frequency between the high-frequency qubit states $|2, -3/2, -\rangle$ and $|0, -1/2, -\rangle$ explored in this work is determined in a spectroscopy sequence in which the molecule is prepared in the state $|2, -3/2, -\rangle$, followed by a comb Raman pulse that transfers the population to $|0, -1/2, -\rangle$ when $|Nf_{\text{rep}} - 2f_{\text{AOM}}|$ is near resonance with the transition (see also ref.³). The molecular population in $|2, -3/2, -\rangle$ is checked after the comb Raman pulse. Absence of population in the state is attributed to populating the $|0, -1/2, -\rangle$ state, because we cannot directly detect this state as in ref.³. We trace out the transition line-shape—that is, the transition probability as a function of f_{AOM} —by repeating the sequence to build up statistics while varying f_{AOM} . To determine the absolute transition frequency, one needs to find the integer N . This is accomplished by measuring the change in frequency Δf_{AOM} that needs to be applied to f_{AOM} to drive the same transition when f_{rep} is changed by Δf_{rep} . Specifically, to maintain $Nf_{\text{rep}} - 2f_{\text{AOM}} = N(f_{\text{rep}} + \Delta f_{\text{rep}}) - 2(f_{\text{AOM}} + \Delta f_{\text{AOM}})$, we arrive at $N = \frac{2\Delta f_{\text{AOM}}}{\Delta f_{\text{rep}}}$. The integer N is determined in this way to be 10,825 for $f_{\text{AOM}} \approx 165.0$ MHz and $f_{\text{rep}} \approx 79.0$ MHz, and $f_{\text{Raman}} \approx 854.8$ GHz. Combined with the experimentally determined rotational constant (~ 142.5 GHz)³, this confirms that the observed transition is between the $J=0$ and $J=2$ rotational manifolds.

Data availability

The data that support the findings of this work are available from the corresponding author upon reasonable request.

Code availability

The computer code used to analyse the data is available from the corresponding author upon reasonable request.

49. Roos, C. F. et al. Experimental demonstration of ground state laser cooling with electromagnetically induced transparency. *Phys. Rev. Lett.* **85**, 5547–5550 (2000).
50. Roos, Ch. et al. Quantum state engineering on an optical transition and decoherence in a Paul trap. *Phys. Rev. Lett.* **83**, 4713–4716 (1999).

Acknowledgements We thank J. Du, J. J. Bollinger and A. L. Collopy for reading and providing feedback on this manuscript, and C. Kurz for help on the experimental setup. This work was supported by the US Army Research Office (grant number W911NF-19-1-0172). Y.L. acknowledges support from the National Key R&D Program of China (grant number 2018YFA0306600), the National Natural Science Foundation of China (grant number 11974330) and Anhui Initiative in Quantum Information Technologies (grant number AHY050000).

Author contributions Y.L., C.-w.C., D.R.L. and D.L. conceived and designed the experiments and contributed to the development of experimental methods and pulse sequences.

C.-w.C., D.R.L. and D.L. developed components of the experimental apparatus. Y.L. collected and analysed the data. Y.L., C.-w.C. and D.L. wrote the manuscript. All authors provided suggestions for the experiments, discussed the results and contributed to the editing of the manuscript.

Competing interests The authors declare no competing interests.

Additional information

Correspondence and requests for materials should be addressed to Y.L.

Peer review information *Nature* thanks Lawrence Cheuk, Barak Dayan and the other, anonymous, reviewer(s) for their contribution to the peer review of this work.

Reprints and permissions information is available at <http://www.nature.com/reprints>.

A biomimetic eye with a hemispherical perovskite nanowire array retina

<https://doi.org/10.1038/s41586-020-2285-x>

Received: 10 April 2019

Accepted: 17 March 2020

Published online: 20 May 2020

 Check for updates

Leilei Gu¹, Swapnadeep Poddar¹, Yuanjing Lin^{1,2}, Zhenghao Long¹, Daquan Zhang¹, Qianpeng Zhang¹, Lei Shu¹, Xiao Qiu¹, Matthew Kam¹, Ali Javey^{2,3} & Zhiyong Fan¹✉

Human eyes possess exceptional image-sensing characteristics such as an extremely wide field of view, high resolution and sensitivity with low aberration¹. Biomimetic eyes with such characteristics are highly desirable, especially in robotics and visual prostheses. However, the spherical shape and the retina of the biological eye pose an enormous fabrication challenge for biomimetic devices^{2,3}. Here we present an electrochemical eye with a hemispherical retina made of a high-density array of nanowires mimicking the photoreceptors on a human retina. The device design has a high degree of structural similarity to a human eye with the potential to achieve high imaging resolution when individual nanowires are electrically addressed. Additionally, we demonstrate the image-sensing function of our biomimetic device by reconstructing the optical patterns projected onto the device. This work may lead to biomimetic photosensing devices that could find use in a wide spectrum of technological applications.

Biological eyes are arguably the most important sensing organ for most of the animals on this planet. In fact, our brains acquire more than 80% of information about our surroundings via our eyes⁴. A human eye with a concavely hemispherical retina and light-management components is particularly notable for its exceptional characteristics including a wide field of view (FOV) of 150°–160°, a high resolution of 1 arcmin per line pair at the fovea and excellent adaptivity to the optical environment¹. Particularly, the domed shape of the retina has the merit of reducing the complexity of optical systems by directly compensating the aberration from the curved focal plane⁵. Mimicking human eyes, artificial vision systems are just as essential in autonomous technologies such as robotics. Particularly for humanoid robots, the vision system should resemble that of a human in appearance to enable amicable human–robot interaction, in addition to having superior device characteristics. In principle, a hemispherical image sensor design mimicking that of the human retina can achieve this goal. However, commercial charge-coupled device (CCD) and complementary-metal-oxide-semiconductor (CMOS) image sensors are mainly using planar device structures shaped by mainstream planar microfabrication processes, making hemispherical device fabrication almost impossible.

Here we demonstrate an artificial visual system using a spherical biomimetic electrochemical eye (EC-EYE) with a hemispherical retina made of a high-density perovskite nanowire array grown using a vapour-phase approach. An ionic liquid electrolyte was used as a front-side common contact to the nanowires and liquid-metal wires were used as back contacts to the nanowire photosensors, mimicking human nerve fibres behind the retina. Device characterizations show that the EC-EYE has a high responsivity, a reasonable response speed, a low detection limit and a wide FOV. The EC-EYE also demonstrates the basic function of a human eye to acquire image patterns. In addition to its structural similarity with a human eye, the hemispherical artificial

retina has a nanowire density much higher than that of photoreceptors in a human retina and can thus potentially achieve higher image resolution, which is bolstered by implementation of a single-nanowire ultrasmall photodetector.

Figure 1 shows a comparison of the human (Fig. 1a–c) and EC-EYE imaging systems (Fig. 1d–f). The human visual system has two eye bulbs for optical sensing, millions of nerve fibres for data transmission and a brain for data processing. The human brain has enormous capability for parallel processing: neuroelectric signals from about a million nerve fibres can be processed simultaneously, enabling high-speed image processing and recognition⁶. The internal structure of a human eye (Fig. 1b) has a lens, a spherical cavity and a hemispherical retina, which is the core component required to convert optical images to neuroelectric signals. Its hemispherical shape simplifies optical design, resulting in an extraordinarily large FOV of about 155° with a wide visual perception of the surroundings¹. There are about 100–120 million photoreceptors and rod and cone cells, vertically assembled in the retina in a dense and quasi-hexangular manner (Fig. 1c), with a density of around 10 million per square centimetre and an average pitch of 3 μm, leading to a high imaging resolution comparable to that of the state-of-the-art CCD/CMOS sensors⁷. However, the nerve fibre layer is at the front surface of the human retina, causing light loss and blind spot issues (Supplementary Fig. 1)¹. Figure 1d, e illustrates the schematic of our biomimetic visual system, which consists of a lens, a photosensor array on a hemispherical substrate and thin liquid-metal wires as electrical contacts. These components mimic the biological eye's lens, retina and the nerve fibres behind the retina, respectively. Of these, the key component is the artificial retina made of a high-density array of perovskite nanowires grown inside a hemispherical porous aluminium oxide membrane (PAM) via a vapour-phase deposition process^{8–10}.

¹Department of Electronic and Computer Engineering, The Hong Kong University of Science and Technology, Kowloon, Hong Kong SAR, China. ²Electrical Engineering and Computer Sciences, University of California, Berkeley, CA, USA. ³Materials Sciences Division, Lawrence Berkeley National Laboratory, Berkeley, CA, USA. ✉e-mail: eezfan@ust.hk

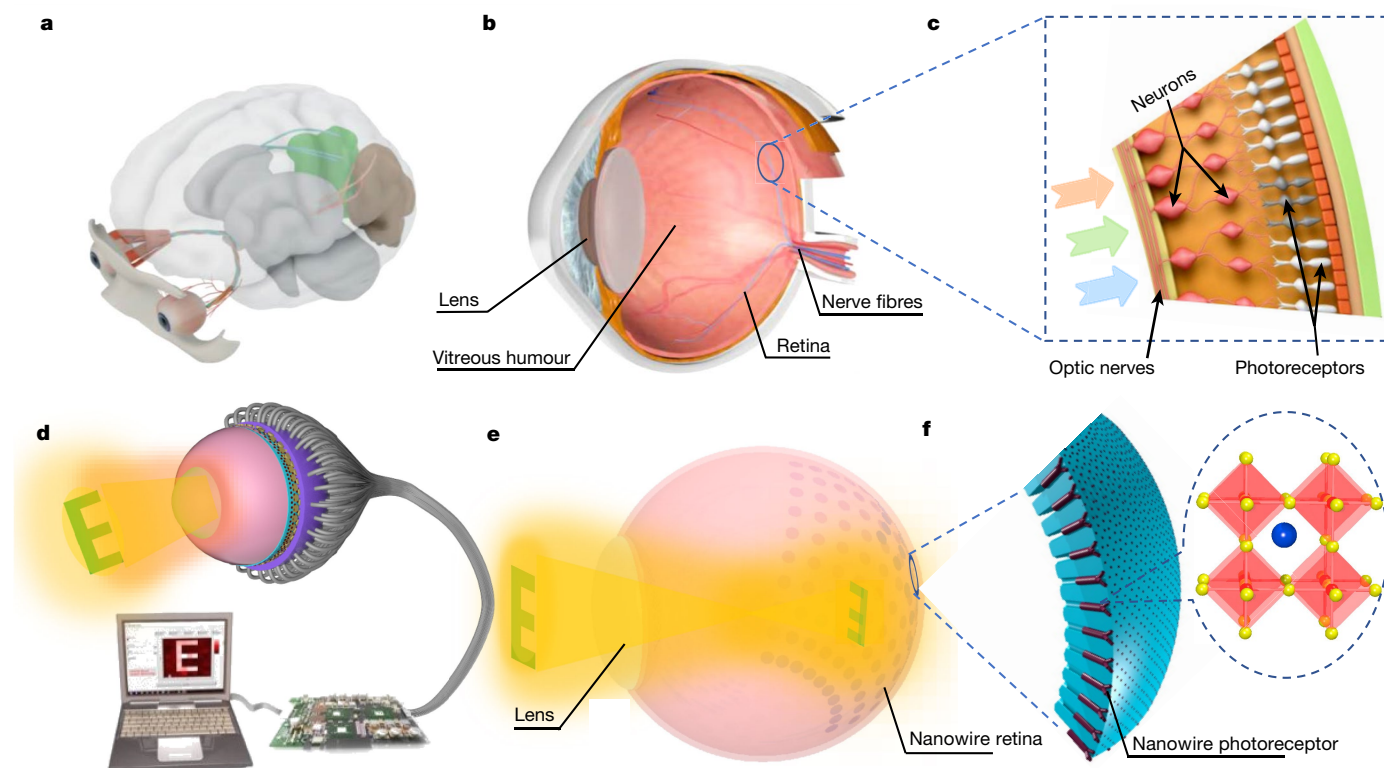


Fig. 1 | Overall comparison of the human visual system and the EC-EYE imaging system. a–c, Schematic of the human visual system (a), the human eye (b) and the retina (c). **d–f,** Schematic of our EC-EYE imaging system (d), the

working mechanism of EC-EYE (e) and perovskite nanowires in the PAM template and their crystal structure (f).

A more detailed structure of the EC-EYE is in Fig. 2a, and the fabrication process is in Methods and Supplementary Fig. 2. The nanowires serve as light-sensitive working electrodes. The tungsten (W) film on aluminium (Al) hemispherical shell works as the counter electrode. In between two electrodes, ionic liquid is used to fill in the cavity, serving as the electrolyte and mimicking the vitreous humour in the human eye. The flexible eutectic gallium indium liquid-metal wires in soft rubber tubes are used for signal transmission between the nanowires and external circuitry with a discontinuous indium layer between the liquid metal and the nanowires to improve contact (Supplementary Fig. 3). An individual photodetector can be addressed and measured by selecting the corresponding liquid-metal wire. This resembles the working principle of the human retina, in which groups of photoreceptors are individually connected with nerve fibres¹¹, enabling suppressed interference among pixels and high-speed parallel processing of the neuroelectric signals. We note that the liquid-metal wires are behind the sensing material, thus avoiding the light-loss and blind-spot problems of the human retina. As a proof of concept, we fabricated a 10×10 photodetector array with a pitch of 1.6 mm. The minimum size of each sensing pixel is limited by the diameter of the liquid-metal wire, which is difficult to reduce to a few micrometres¹². To further reduce pixel size and enhance the spatial imaging resolution, another approach to fabricating the sensor pixel array with a pixel area of about $1 \mu\text{m}^2$ per pixel using metal microneedles has been developed.

Previously, there have been a few inspiring works reporting hemispherical image sensors using deformed, folded or individually assembled photodetectors^{13–15}. The photodetectors in those works were mainly pre-fabricated on planar substrates, then transferred to a hemispherical supporting material or folded into a hemispherical shape. It is challenging to achieve small individual pixel size and high imaging resolution owing to the complexity of the fabrication process. Here light-sensing nanowires were grown in a hemispherical template, and

thus a structure akin to that of the human retina has been formed in a single step. We chose formamidinium lead iodide (FAPbI₃) as the model material for nanowire growth here owing to its excellent optoelectronic properties and decent stability^{9,16}. The nanowire growth and characterization details can be found in Methods and Supplementary Figs. 4 and 5. In principle, other types of inorganic nanowires made of Si, Ge, GaAs and so on can also be grown using the well documented vapour–liquid–solid process^{17–20}. Figure 2b, c shows the side and top views of a completed EC-EYE. Figure 2d, e presents scanning electron microscopy (SEM) images of the hemispherical PAM and the nanowires located at the bottom of the nanochannels, respectively. The single-crystalline (Fig. 2f) nanowires have a pitch of 500 nm, corresponding to a density of $4.6 \times 10^8 \text{ cm}^{-2}$, which is much higher than that of the photoreceptors in human retina, indicating the potential to achieve a high imaging resolution if proper electrical contacts can be achieved²¹.

Figure 3a shows the schematic of a single pixel measurement. A collimated light beam is focused on the pixels at the centre of the retina. Figure 3b plots the energy-band alignment for the entire device showing charge-carrier separation routes under light excitation. Figure 3c presents the current–voltage characteristics exhibiting the asymmetric photoresponse caused by asymmetric charge transportation at the two sides of the nanowires (Fig. 3b). Previous electrochemical characterizations have shown that the redox reactions of I^-/I_3^- pairs^{22,23} occur at the nanowire/electrolyte and tungsten film/electrolyte interfaces and that ion transportation inside the electrolyte contributes to the photoresponse (Supplementary Fig. 6). The inset of Fig. 3c shows the transient response of the device to chopped light (see Supplementary Figs. 7 and 8 for more results). The relatively fast and highly repeatable response indicates that the device has excellent photocurrent stability and reproducibility. The response and recovery time are found to be 32.0 ms and 40.8 ms, respectively. Further electrochemical analysis of the critical nanowire/electrolyte interface reveals that the device

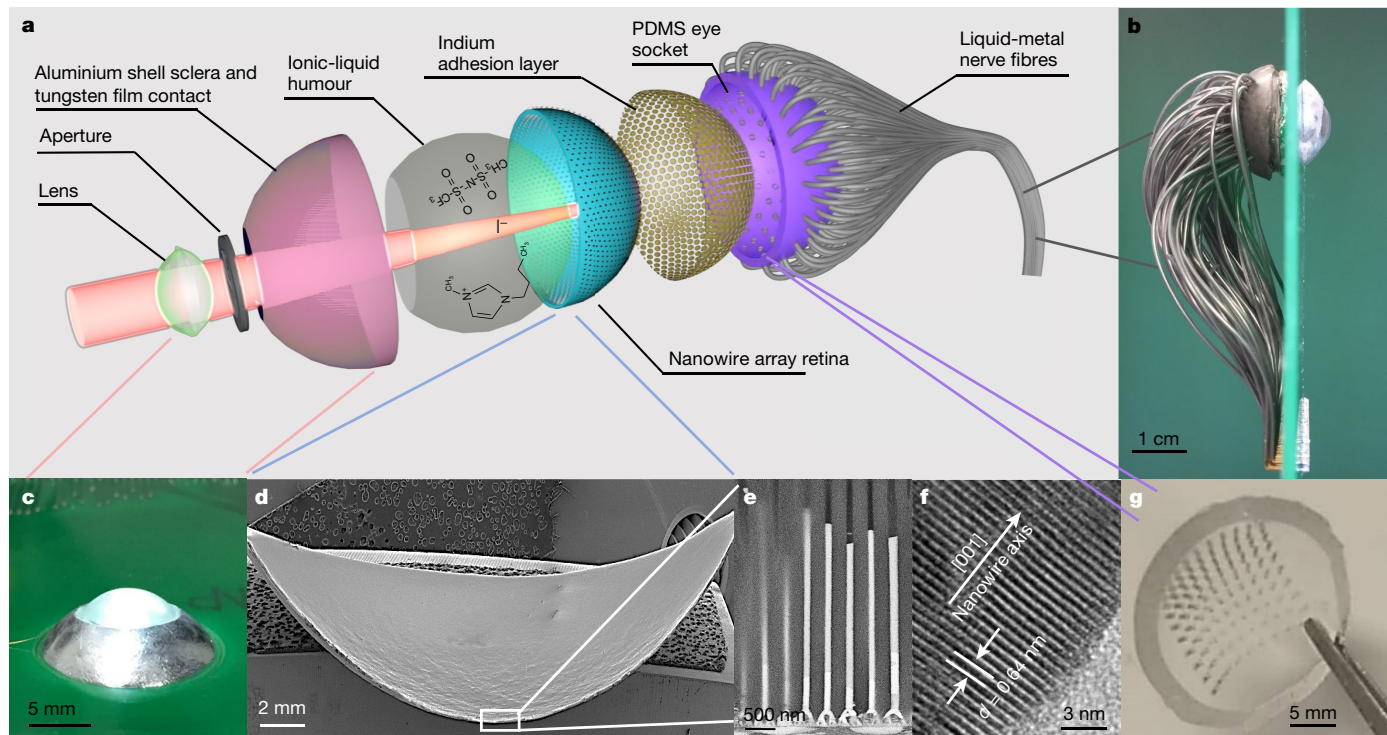


Fig. 2 | Detailed structure of our EC-EYE. a, Exploded view of EC-EYE. **b, c**, Side view (**b**) and top view (**c**) of a completed EC-EYE. **d**, Low-resolution cross-sectional SEM image of the hemispherical PAM/nanowires. **e**, Cross-sectional SEM images of nanowires in PAM. **f**, High-resolution

transmission electron microscopy image of a single-crystalline perovskite nanowire. **g**, Photograph of the polydimethylsiloxane (PDMS) socket, which improves the alignment of the liquid-metal wires.

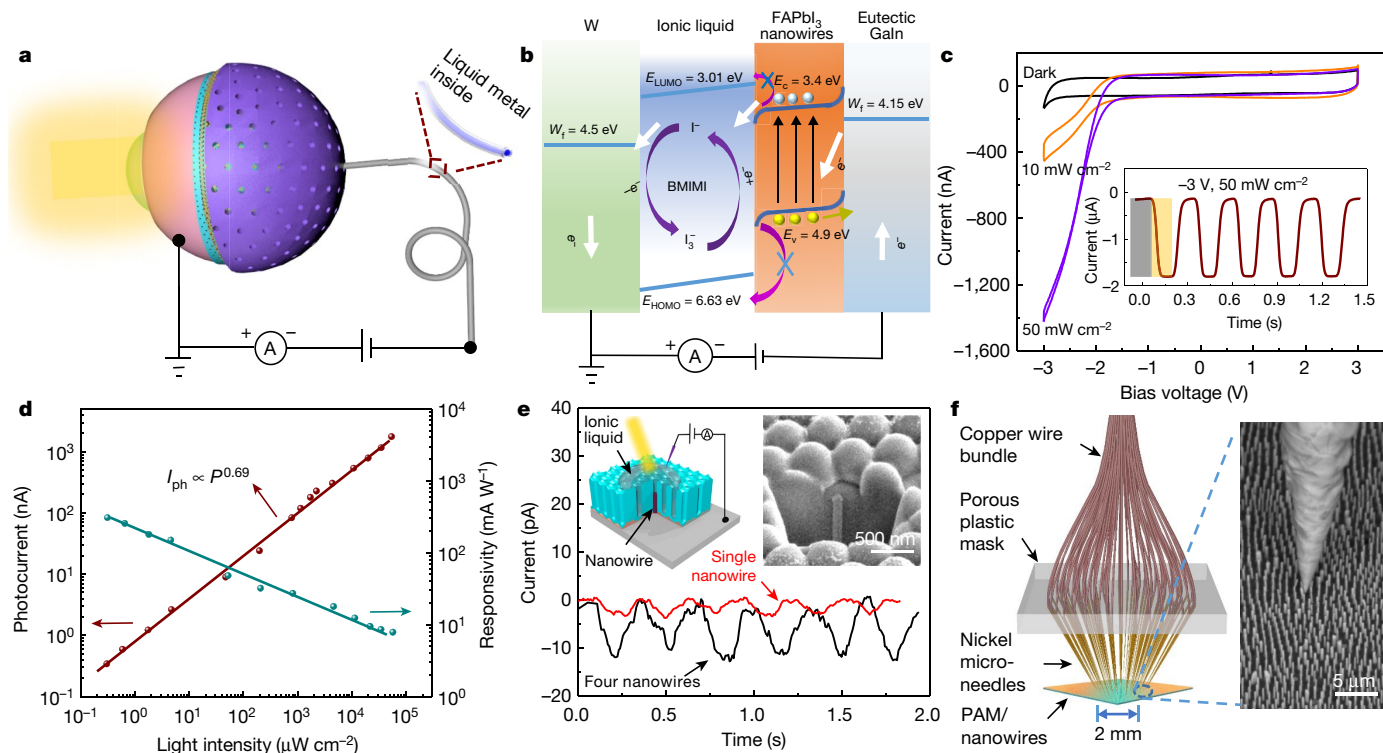


Fig. 3 | Photodetection performance characterization of individual pixels.

a, Schematic setup of individual pixel measurement. **b**, Working mechanism of an individual pixel under -3 V bias voltage. BMIMI, 1-butyl-3-methylimidazolium iodide. **c**, Current-voltage curves under different illuminations and the transient response of individual pixels under the illumination of simulated sunlight with an intensity of 50 mW cm^{-2} . The

current-voltage curves represent one cycle of the cyclic-voltammetry measurement. Scan rate, 100 mV s^{-1} . **d**, Illumination-intensity-dependent photocurrent and responsivity of an individual pixel. The lowest light intensity is $0.3 \mu\text{W cm}^{-2}$. **e**, Device schematic and transient photoresponse of single-nanowire-based and four-nanowire-based individual pixels. **f**, Schematic and SEM image of the Ni microneedle contact to the nanowire array.

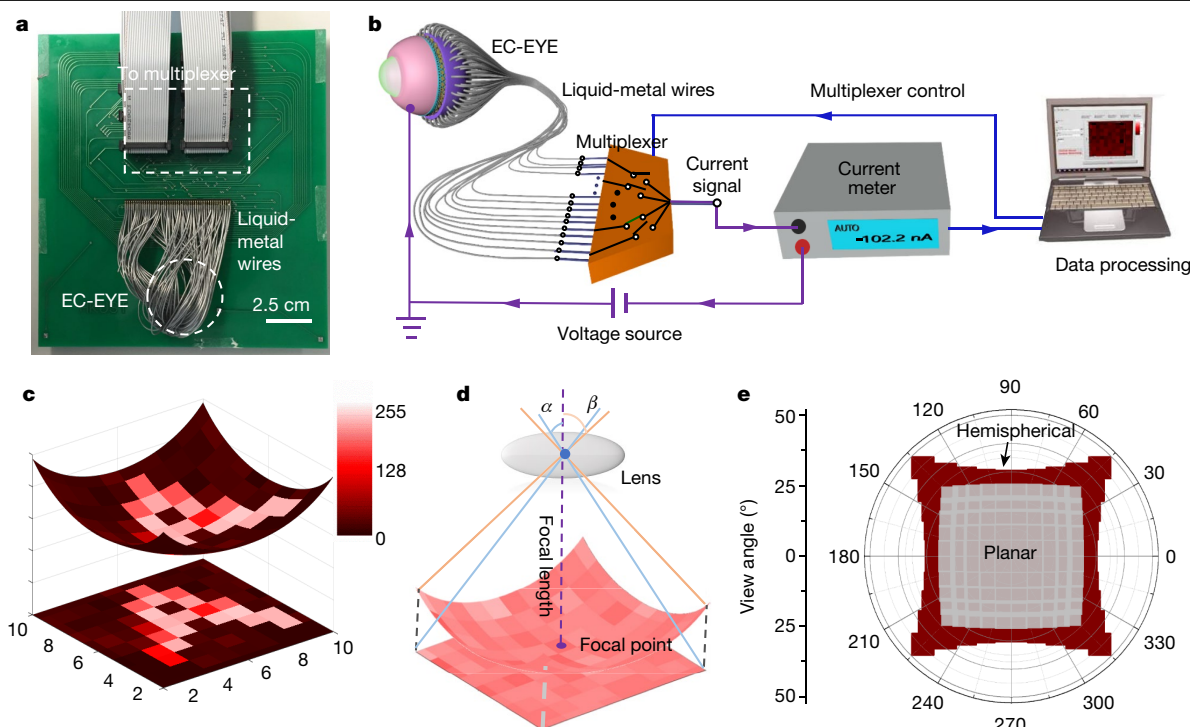


Fig. 4 | Image-sensing demonstration of EC-EYE. **a**, Back-view photo of an EC-EYE mounted on a printed circuit board. **b**, Schematic illustration of the measurement setup. **c**, The reconstructed image (letter 'A') of EC-EYE and its

projection on a flat plane. **d**, **e**, The schematic (**d**) and calculated (**e**) FOV of the planar and hemispherical image-sensing systems.

response time depends on the kinetics of multiple types of ions at that interface (Supplementary Fig. 9). Electrochemical impedance spectroscopy measurements (Supplementary Fig. 10) demonstrate that device structural optimization and ionic liquid concentration increase can substantially reduce the charge-transfer resistance (R_{ct}) at the nanowire/electrolyte interface, leading to reduction of the device response and recovery times to 19.2 ms and 23.9 ms. This is much faster than that of human eyes, whose response and recovery times range from 40 ms to 150 ms (ref. ²⁴). Meanwhile, increasing ionic liquid concentration leads to light absorption loss (Supplementary Fig. 11), so further optimization of the ionic liquid composition will be of benefit.

Figure 3d shows the dependence of the photocurrent and responsivity on the illumination intensity, with a large dynamic range from $0.3 \mu\text{W cm}^{-2}$ to 50 mW cm^{-2} . The photocurrent can be fitted with a quasi-linear power-law relationship ($I = A \times P^{0.69}$), where I is the photocurrent, P represents the irradiance power and A is a coefficient. Intriguingly, the responsivity increases when reducing illumination intensity. It can reach up to 303.2 mA W^{-1} , which is among the highest for reported photoelectrochemical photodetectors (Supplementary Table 1). And it is on par with that of the solid-state photodetectors based on perovskite nanowire arrays reported earlier^{8,9}. Under the lowest radiation level measured, the average number of photons received per second by an individual nanowire can be estimated at 86 photons (Supplementary Information). This sensitivity is on par with that of human cone cells²⁵. The corresponding specific detectivity is calculated as about 1.1×10^9 Jones for $0.3 \mu\text{W cm}^{-2}$ incident light. The spectral responsivity shows a broad-band response with a clear cut-off at 810 nm (Supplementary Fig. 12). Supplementary Fig. 13 demonstrates the stability and repeatability of an individual pixel for 2 Hz light continuously chopped for 9 h. It indicates that although there are drifts for both the dark and light current, there is no obvious device performance degradation after 64,800 cycles.

As mentioned, one of the primary merits of using high-density nanowire arrays for artificial retina is their potential for high image

resolution. Although liquid-metal fibre contacts to nanowires are convenient and image resolution is already on par with that of a number of existing bionic eyes in use^{26–28}, it is challenging to reduce pixel size down to the few-micrometres level. Therefore, we explored two more contact strategies to achieve ultrasmall pixel size. As shown in Fig. 3e, a single nanowire was deterministically grown in a single nanochannel opened using a focused ion beam; then a single pixel with 500-nm lateral size and a footprint of about $0.22 \mu\text{m}^2$ was achieved (Methods and Supplementary Fig. 14). Using the same approach, a pixel of 4 nanowires was also fabricated. The SEM images in Supplementary Fig. 15 show the controllable growth of nanowires, including the nanowire numbers and positions. The photoresponses of these two devices are shown in Fig. 3e. To form an array of ultrasmall pixels, nickel (Ni) microneedles were vertically assembled on top of a PAM using a magnetic field and thus each microneedle can address 3 nanowires, forming a pixel with lateral size of about $1 \mu\text{m}$ and pitch of $200 \mu\text{m}$. The details of this contact strategy are illustrated in Supplementary Figs. 16 and 17. Figure 3f schematically shows the device connected to copper wires, which serve as signal transmission lines. The lateral size of the contact region is 2 mm.

After characterization of individual sensor pixels, we measured the full device imaging functionality. A photograph of the device is shown in Fig. 4a and Supplementary Fig. 18a, b. The liquid-metal wires are connected to a computer-controlled 100×1 multiplexer via a printed circuit board. The measurement system design is shown in Fig. 4b and the corresponding circuit diagram is shown in Supplementary Fig. 18c. The image-sensing function was examined by projecting optical patterns onto the EC-EYE, after which the photocurrent of each sensor pixel was recorded. Before pattern generation and recognition, the consistency of the dark and light currents of all pixels was verified. Supplementary Fig. 19a, b depicts the dark- and light-current images obtained under -3 V bias voltage, showing that all 100 pixels have consistent photoresponse with relatively small variation. To reconstruct the optical pattern projected on the EC-EYE, a photocurrent value was converted to a greyscale number between 0 and 255 (Supplementary Information).

Figure 4c shows the imaged character 'A' and its projection onto a flat plane. Supplementary Fig. 20 shows images of letters 'E', 'I' and 'Y'. The Supplementary Video shows the dynamic process of EC-EYE capturing the letters 'E', 'Y' and 'E' sequentially. Compared to planar image sensors based on a crossbar structure, the device presented here delivers a higher contrast with clearer edges because each individual pixel is better isolated from the neighbouring pixels (Supplementary Fig. 21)⁷. Besides an EC-EYE with liquid-metal contacts, we also fabricated a small electrochemical image sensor with microneedle contacts. This imager, with an active area of 2 mm × 2 mm, was assembled into a mini-camera together with other optical parts (see Supplementary Figs. 22 and 23). Supplementary Fig. 23 shows its imaging functionality. Meanwhile, the magnetic microneedle alignment strategy developed here also works very well for the whole hemispherical surface. Supplementary Fig. 24 shows that the 50-μm-thick Ni microwires are well aligned onto the surface of a hemispherical PAM. Although we have successfully fabricated ultrasml pixels here and implemented microneedle manual alignment onto the nanowires, better high-throughput strategies to align large numbers of electrodes on nanowires with precision could be developed. For instance, a high-precision robotic arm equipped with a piezo actuator can be used to place Ni microneedles onto the hemispherical PAM, assisted by a magnetic field and a high-resolution optical monitoring system (Supplementary Fig. 24c). Better approaches could address individual nanowires in a more deterministic manner.

Compared to a planar image sensor, the hemispherical shape of our EC-EYE ensures a more consistent distance between pixels and the lens, resulting in a wider FOV and better focusing onto each pixels (Fig. 4d). The diagonal visual field of our hemispherical EC-EYE is about 100.1°, whereas that of a planar device is only 69.8° (Fig. 4e). Moreover, this angle of view can be further improved to approach the static vertical FOV of a single human eye (about 130°)²⁹, by optimizing the pixel distribution on the hemispherical retina.

Here we have demonstrated a biomimetic eye with a hemispherical retina made of high-density light-sensitive nanowires. Its structure has a high degree of similarity to that of a human eye with potential to achieve higher imaging resolution if a better contact strategy can be implemented. The processes developed tackle the challenge of fabricating optoelectronic devices on non-planar substrates with high integration density. Furthermore, this work may inspire biomimetic designs of optical imaging devices that could find application in scientific instrumentation, consumer electronics and robotics.

Online content

Any methods, additional references, Nature Research reporting summaries, source data, extended data, supplementary information, acknowledgements, peer review information; details of author contributions and competing interests; and statements of data and code availability are available at <https://doi.org/10.1038/s41586-020-2285-x>.

- Atchison, D. A. & Smith, G. *Optics of the Human Eye* Vol. 35 (Butterworth-Heinemann, 2000).
- Zhang, J., Con, C. & Cui, B. Electron beam lithography on irregular surfaces using an evaporated resist. *ACS Nano* **8**, 3483–3489 (2014).
- Qin, D., Xia, Y. & Whitesides, G. M. Soft lithography for micro-and nanoscale patterning. *Nat. Protocols* **5**, 491–502 (2010).
- Pocock, D. C. D. Sight and knowledge. *Trans. Inst. Br. Geogr.* **6**, 385–393 (1981).
- Jung, I. et al. Dynamically tunable hemispherical electronic eye camera system with adjustable zoom capability. *Proc. Natl Acad. Sci. USA* **108**, 1788–1793 (2011).
- Nassi, J. J. & Callaway, E. M. Parallel processing strategies of the primate visual system. *Nat. Rev. Neurosci.* **10**, 360–372 (2009).
- Jonas, J. B., Schneider, U. & Naumann, G. O. Count and density of human retinal photoreceptors. *Graefes Arch. Clin. Exp. Ophthalmol.* **230**, 505–510 (1992).
- Gu, L. et al. 3D arrays of 1024-pixel image sensors based on lead halide perovskite nanowires. *Adv. Mater.* **28**, 9713–9721 (2016).
- Gu, L. et al. Significantly improved black phase stability of FAPbI₃ nanowires via spatially confined vapor phase growth in nanoporous templates. *Nanoscale* **10**, 15164–15172 (2018).
- Waleed, A. et al. Lead-free perovskite nanowire array photodetectors with drastically improved stability in nanoengineering templates. *Nano Lett.* **17**, 523–530 (2017).
- Schein, S. J. Anatomy of macaque fovea and spatial densities of neurons in foveal representation. *J. Comp. Neurol.* **269**, 479–505 (1988).
- Dickey, M. D. et al. Eutectic gallium-indium (EGaIn): a liquid metal alloy for the formation of stable structures in microchannels at room temperature. *Adv. Funct. Mater.* **18**, 1097–1104 (2008).
- Song, Y. M. et al. Digital cameras with designs inspired by the arthropod eye. *Nature* **497**, 95–99 (2013).
- Ko, H. C., Stoykovich, M. P., Song, J., Malyarchuk, V. & Rogers, J. A. A hemispherical electronic eye camera based on compressible silicon optoelectronics. *Nature* **454**, 748–753 (2008).
- Zhang, K. et al. Origami silicon optoelectronics for hemispherical electronic eye systems. *Nat. Commun.* **8**, 1782 (2017).
- Han, Q. et al. Single crystal formamidinium lead iodide (FAPbI₃): insight into the structural, optical, and electrical properties. *Adv. Mater.* **28**, 2253–2258 (2016).
- Fan, Z. et al. Ordered arrays of dual-diameter nanopillars for maximized optical absorption. *Nano Lett.* **10**, 3823–3827 (2010).
- Fan, Z. et al. Three-dimensional nanopillar-array photovoltaics on low-cost and flexible substrates. *Nat. Mater.* **8**, 648–653 (2009).
- Ramdani, M. R. et al. Fast growth synthesis of GaAs nanowires with exceptional length. *Nano Lett.* **10**, 1836–1841 (2010).
- Wen, C. Y. et al. Formation of compositionally abrupt axial heterojunctions in silicon-germanium nanowires. *Science* **326**, 1247–1250 (2009).
- Wandell, B. A. *Foundations of Vision* Vol. 8 (Sinauer Associates, 1995).
- Boschloo, G. & Hagfeldt, A. Characteristics of the iodide/triiodide redox mediator in dye-sensitized solar cells. *Acc. Chem. Res.* **42**, 1819–1826 (2009).
- Kawano, R. & Watanabe, M. Equilibrium potentials and charge transport of an I[−]/I₃[−] redox couple in an ionic liquid. *Chem. Commun.* **3**, 330–331 (2003).
- Rayner, K., Smith, T. J., Malcolm, G. L. & Henderson, J. M. Eye movements and visual encoding during scene perception. *Psychol. Sci.* **20**, 6–10 (2009).
- Mustafi, D., Engel, A. H. & Palczewski, K. Structure of cone photoreceptors. *Prog. Retin. Eye Res.* **28**, 289–302 (2009).
- Fujikado, T. et al. Evaluation of phosphenes elicited by extraocular stimulation in normals and by suprachoroidal-transretinal stimulation in patients with retinitis pigmentosa. *Graefes Arch. Clin. Exp. Ophthalmol.* **245**, 1411–1419 (2007).
- Ayton, L. N. et al. First-in-human trial of a novel suprachoroidal retinal prosthesis. *PLoS One* **9**, e115239 (2014).
- Shivdasani, M. N. et al. Factors affecting perceptual thresholds in a suprachoroidal retinal prosthesis. *Invest. Ophthalmol. Vis. Sci.* **55**, 6467–6481 (2014).
- Navarro, R. The optical design of the human eye: a critical review. *J. Optom.* **2**, 3–18 (2009).

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

© The Author(s), under exclusive licence to Springer Nature Limited 2020

Methods

Fabrication of EC-EYE

The EC-EYE fabrication process started with deforming a 500- μm -thick Al sheet on a set of hemispherical moulds to obtain a hemispherical Al shell, which then underwent a standard two-step anodization process to form PAM with thickness 40 μm and nanochannel pitch and diameter of 500 nm and 120 nm, respectively, on the Al surface. A barrier thinning process and Pb electrodeposition were carried out to obtain Pb nanoclusters at the bottom of the PAM channels. Next, the outer layer of PAM and the residual Al were etched away to obtain a freestanding PAM with Pb, which was then transferred into a tubular furnace to grow perovskite nanowires about 5 μm long. The detailed nanowire growth condition can be found in ref.⁹ A 20-nm-thick indium layer was evaporated onto the PAM back surface to serve as the adhesion layer. We note that this indium layer will not cause short-circuiting between pixels owing to its discontinuous morphology (Supplementary Fig. 3). To obtain the liquid-metal contact array, a hedgehog-shaped mould was fabricated using 3D printing, from which a complementary PDMS socket with 10×10 hole array (hole size 700 μm , pitch 1.6 mm) was cast. Eutectic gallium indium liquid metal was then injected into thin soft tubes (inner diameter 400 μm , outer diameter 700 μm) to form liquid-metal wires. Then 100 tubes were inserted into the holes on the PDMS socket and the whole socket was attached to the PAM/nanowire surface to form a 10×10 photodetector array. These long soft tubes can be directly connected to a printed circuit board and thus the complex wire bonding process is avoided. A circular hole was opened on another Al shell, which was then coated with a tungsten film working as the counter electrode of the EC-EYE. After mounting the aperture (Eakins, SK6), the Al shell was subsequently fixed onto the front side of PAM by epoxy. Ionic liquid 1-butyl-3-methylimidazolium bis(trifluoromethylsulfonyl)imide (BMIMTFSI) mixed with 1 vol% 1-butyl-3-methylimidazolium iodide (BMIMI) was then injected and a convex lens (diameter 1.2 cm, focal length 1.6 cm) was then glued to the hole on Al shell to seal the device. After curing, the EC-EYE device fabrication was completed.

Fabrication of microneedle-based electrochemical image sensor

Freestanding 40- μm -thick PAM was fabricated using the standard anodization process, NaOH etching and HgCl_2 solution etching. Ion milling was used to remove the barrier layer to achieve through-hole PAM (Supplementary Fig. 17a). Then a 1- μm -thick Cu film was thermally evaporated onto the through-hole PAM to serve as the electrode for the subsequent Ni and Pb electrochemical deposition (Supplementary Fig. 17b). Next, to expose the Ni nanowires, the copper layer was removed by Ar^+ ion milling and the PAM was partially etched away by reactive ion etching. The exposed Ni nanowires were about 3 μm long (Supplementary Fig. 17c). The chip was moved into a tubular furnace for perovskite nanowire growth (Supplementary Fig. 17c). The PAM chip was fixed onto a cylindrical electromagnet (0–50 mT) with Ni nanowires facing upward (Supplementary Fig. 17d). Meanwhile, Ni microwires of diameter 50 μm were sharpened in a mixed acidic solution (100 ml 0.25 M HCl aqueous solution + 100 ml ethylene glycol) under a bias of 1 V, with Ni microwires as the working electrodes and the tungsten coil as the counter electrode. The resulting Ni microwires have sharp tips, with curvature radius of 100–200 nm. The Ni needle was then gently placed onto the PAM substrate with the magnetic field 'on'. Owing to the magnetic force, the ferromagnetic Ni microneedles can engage into the Ni nanowire forest to form an effective electrical contact to the nanowires (Supplementary Fig. 17d). To facilitate the Ni microwire placement, a mask with 10×10 hole array (hole diameter 100 μm , pitch 200 μm) was used to align the Ni microneedles (Supplementary Fig. 17d). After placement, ultraviolet epoxy was dropped between the mask and the PAM substrate. Copper enamelled wire with diameter of 60 μm was inserted into the hole to form an electrical contact bridging the Ni microneedle and external printed circuit board (Supplementary Fig. 17e).

Fabrication of single- and multiple-nanowire-based electrochemical photodetectors

Freestanding planar PAM was fabricated by standard two-step anodization followed by HgCl_2 etching. The freestanding PAM was then transferred into the focused ion beam (FEI Helios G4 UX) to selectively etch away the barrier layer (Supplementary Fig. 14c). To facilitate the etching, the chip was bonded onto an Al substrate with the barrier layer side facing up. After focused-ion-beam etching (etching voltage 30 kV, etching current 26 nA), a 500-nm-thick Cu layer was evaporated onto the barrier layer side to serve as the electrode for the subsequent Pb electrochemical deposition (Supplementary Fig. 14d). Next, the chip was moved into a tubular furnace for perovskite nanowire growth (Supplementary Fig. 14e). Then a Cu wire was bonded onto the Cu side of PAM with carbon paste and the whole chip was fixed onto a glass substrate by ultraviolet epoxy. After curing, ionic liquid was dropped onto the top of the PAM (Supplementary Fig. 14e) and a tungsten probe was inserted into the ionic liquid for photoelectric measurement. The photoresponse was measured with -3 V bias and 50 mW cm^{-2} light intensity.

Material and photodetector characterization

The SEM images and energy dispersive X-ray mapping of the PAM were characterized using a field-emission scanning electron microscope (JEOL JSM-7100F equipped with a Si (Li) detector and PGT 4000T analyser). The X-ray diffraction patterns of the FAPbI_3 nanowire arrays in PAM were obtained using a Bruker D8 X-ray diffractometer. Transmission electron microscope images were obtained using a TEM JEOL (2010) with 200-kV acceleration voltage. The ultraviolet–visible absorption was measured with a Varian Cary 500 spectrometer. The photoluminescence and time-resolved photoluminescence measurements were carried out on an Edinburgh F55 fluorescence spectrometer. The cyclic-voltammetry measurements based on a two-electrode configuration were performed on an electrochemical workstation (CHI 660E, China) at a scan rate of 100 mV s^{-1} with attenuated simulated sunlight (Newport, Solar Spectral Irradiance Air Mass 1.5) as the light source. The current–time curves of individual pixels were measured using the probe station of a HP4156A with neutral-density filters to tune the light intensity. An additional chopper was used to chop light into square wave optical signals with different frequencies. The electrochemical impedance spectra were measured by a potentiostat (Gamry SG 300) in the range 300 kHz to 100 Hz, with an amplitude of 10 mV under a bias of -3 V. The working electrodes were connected to the liquid metal. The reference and counter electrodes were connected to the tungsten electrode.

Image-sensing characterization of EC-EYE

The image-sensing performance of EC-EYE was characterized by using a home-built system consisting of a multiplexer, a pre-amplifier, a laptop computer and a Labview program (<https://www.ni.com/zh-cn/shop/labview.html>). The schematic of the system can be found in Fig. 4b. Specifically, Keithley 2400 was used to provide the bias voltage. The current meter (PXI4130, National Instruments), together with the multiplexer (PXI2530B, National Instruments), was installed inside of a chassis (PXI1031, National Instruments). The whole system is controlled by a home-built Labview program. To carry out the measurements, various optical patterns were generated by PowerPoint slides and projected onto the device by a projector. A convex lens was used to focus the pattern and different neutral-density filters were inserted between the projector and image sensor to tune the light intensity.

Reporting summary

Further information on research design is available in the Nature Research Reporting Summary linked to this paper.

Data availability

The data that support the findings of this study are provided in the main text and the Supplementary Information. More data are available from the corresponding author upon reasonable request.

Acknowledgements This work was supported by the National Natural Science Foundation of China (project 51672231) and the Science and Technology Plan of Shenzhen (JCYJ20170818114107730), the General Research Fund (projects 16237816, 16309018 and 16214619) from the Hong Kong Research Grant Council, Hong Kong Innovation Technology Commission (project ITS/115/18) and The Hong Kong University of Science and Technology (HKUST) Fund of Nanhai (grant number FSNH-18FYTRI01). We acknowledge support received from the Material Characterization and Preparation Facility (MCPF, particularly Y. Cai for her technical assistance with the focused ion beam), the Nanosystem Fabrication Facility (NFF), the Center for 1D/2D Quantum Materials and the State Key Laboratory on Advanced Displays and Optoelectronics at HKUST. We also thank D. Wang (Chemistry Department, Boston College), M. Shao (Department of Chemical and Biological Engineering, HKUST) and Q. Chen (Department of Mechanical and Aerospace Engineering, HKUST) for their discussions on the

electrochemical impedance spectroscopy measurements. The images in Fig. 1a–d and Supplementary Fig. 1 were created by Fantastic Color Animation Technology Co., Ltd (2020).

Author contributions Z.F. and L.G. conceived the ideas of the work. L.G., Z.L., D.Z., Q.Z. and L.S. contributed to PAM fabrication, perovskite and Ni nanowire growth. S.P. contributed to the focused-ion-beam process. Y.L. developed the photoelectrochemical working mechanism. L.G., Z.L. and X.Q. worked on Ni microneedle assembly and EC-EYE device characterizations. L.G., Y.L., M.K., A.J. and Z.F. carried out the data analysis and wrote the manuscript. All authors discussed the results and commented on the manuscript.

Competing interests The authors declare no competing interests.

Additional information

Supplementary information is available for this paper at <https://doi.org/10.1038/s41586-020-2285-x>.

Correspondence and requests for materials should be addressed to Z.F.

Peer review information *Nature* thanks Dae-Hyeong Kim, Hongrui Jiang and the other, anonymous, reviewer(s) for their contribution to the peer review of this work.

Reprints and permissions information is available at <http://www.nature.com/reprints>.

Short-range order and its impact on the CrCoNi medium-entropy alloy

<https://doi.org/10.1038/s41586-020-2275-z>

Received: 31 July 2019

Accepted: 9 March 2020

Published online: 20 May 2020

 Check for updates

Ruopeng Zhang^{1,2,5}, Shiteng Zhao^{1,2,5}, Jun Ding³, Yan Chong^{1,2}, Tao Jia⁴, Colin Ophus², Mark Asta^{1,3}, Robert O. Ritchie^{1,3} & Andrew M. Minor^{1,2}✉

Traditional metallic alloys are mixtures of elements in which the atoms of minority species tend to be distributed randomly if they are below their solubility limit, or to form secondary phases if they are above it. The concept of multiple-principal-element alloys has recently expanded this view, as these materials are single-phase solid solutions of generally equiatomic mixtures of metallic elements. This group of materials has received much interest owing to their enhanced mechanical properties^{1–5}. They are usually called medium-entropy alloys in ternary systems and high-entropy alloys in quaternary or quinary systems, alluding to their high degree of configurational entropy. However, the question has remained as to how random these solid solutions actually are, with the influence of short-range order being suggested in computational simulations but not seen experimentally^{6,7}. Here we report the observation, using energy-filtered transmission electron microscopy, of structural features attributable to short-range order in the CrCoNi medium-entropy alloy. Increasing amounts of such order give rise to both higher stacking-fault energy and hardness. These findings suggest that the degree of local ordering at the nanometre scale can be tailored through thermomechanical processing, providing a new avenue for tuning the mechanical properties of medium- and high-entropy alloys.

Among the increasing number of medium- to high-entropy alloy systems reported in the literature^{8–12}, the CrCoNi-based, face-centred-cubic (fcc) single-phase alloys exhibit an exceptional combination of mechanical properties, including high strength, tensile ductility, fracture toughness and impact resistance¹³. Extensive studies have documented the deformation mechanisms in these alloys. Gludovatz et al. reported the outstanding fracture toughness of CrCoNi at cryogenic temperatures¹⁴, and attributed this to a synergy of deformation mechanisms, including a propensity for mechanical twinning¹⁵. Interestingly, computational work has suggested that the CrCoNi-based fcc single-phase alloys should have near-zero or negative stacking-fault energies (SFEs; γ_{SF})^{15–19}. However, these computational predictions do not agree with measured values^{20,21} ($\gamma_{\text{SF, CrCoNi}} \approx 22 \text{ mJ m}^{-2}$ and $\gamma_{\text{SF, CrMnFeCoNi}} \approx 30 \text{ mJ m}^{-2}$). Experimentally, the measured SFEs in medium-entropy alloys (MEAs) and high-entropy alloys (HEAs) exhibit a wide distribution²², indicating a strong dependence of γ_{SF} on local atomic configuration. Ding et al.⁶ showed that the SFE of CrCoNi MEA can be tailored over a wide range by tuning its local chemical order. The work highlights the potentially strong impact of chemical short-range order (SRO) on the mechanical properties of the MEA/HEAs. Later, Li et al.⁷, using molecular dynamics simulations, demonstrated the ruggedness of the local energy landscape and how it raises activation barriers governing dislocation activities. Experimental evidence for the existence of such SRO has so far been limited to X-ray adsorption measurements²³ that are averaged over a relatively large volume of material. Indeed, further efforts are

needed to characterize the degree and the spatial extent of the ordering, as well as how both would be affected by thermal history and any associated effects on mechanical behaviour. Here we provide quantitative visualization of the SRO structure, by which we establish a direct effect of this SRO on the mechanical behaviour of MEA/HEA materials.

To investigate the presence of chemical SRO, samples of equiatomic CrCoNi alloys were subjected to different thermal treatments after homogenization at 1,200 °C: (1) water-quenched to room temperature to suppress SRO formation; or (2) aged at 1,000 °C for 120 h followed by slow furnace cooling to promote SRO formation. The microstructure and the degree of SRO were characterized with a variety of transmission electron microscope (TEM) imaging techniques. Diffraction contrast from SRO is inherently faint as compared to the fcc matrix lattice diffraction signal because the former arises from relatively minor differences in lattice distortion. As a result, measurement of the faint SRO diffraction signal has proven to be challenging. In order to enhance the signal-to-noise ratio of the diffraction contrast from SRO, we minimized the background noise from inelastic scattering by using a Zeiss TEM (LIBRA 200MC) equipped with an in-column Ω energy filter and a camera with 16-bit dynamic range. Energy-filtered diffraction patterns and dark-field images for the two heat treatment conditions are shown in Fig. 1. In the diffraction patterns (Fig. 1a, b), streaks along $\{111\}$ directions between fcc Bragg spots are clearly observed in the aged sample. Dark-field imaging taken with the objective aperture positioned in the centre of the streaked region shown in Fig. 1b was used to image the

¹Department of Materials Science and Engineering, University of California, Berkeley, CA, USA. ²National Center for Electron Microscopy, Molecular Foundry, Lawrence Berkeley National Laboratory, Berkeley, CA, USA. ³Materials Sciences Division, Lawrence Berkeley National Laboratory, Berkeley, CA, USA. ⁴Department of Physics, Stanford University, Stanford, CA, USA.

⁵These authors contributed equally: Ruopeng Zhang, Shiteng Zhao. ✉e-mail: aminor@lbl.gov

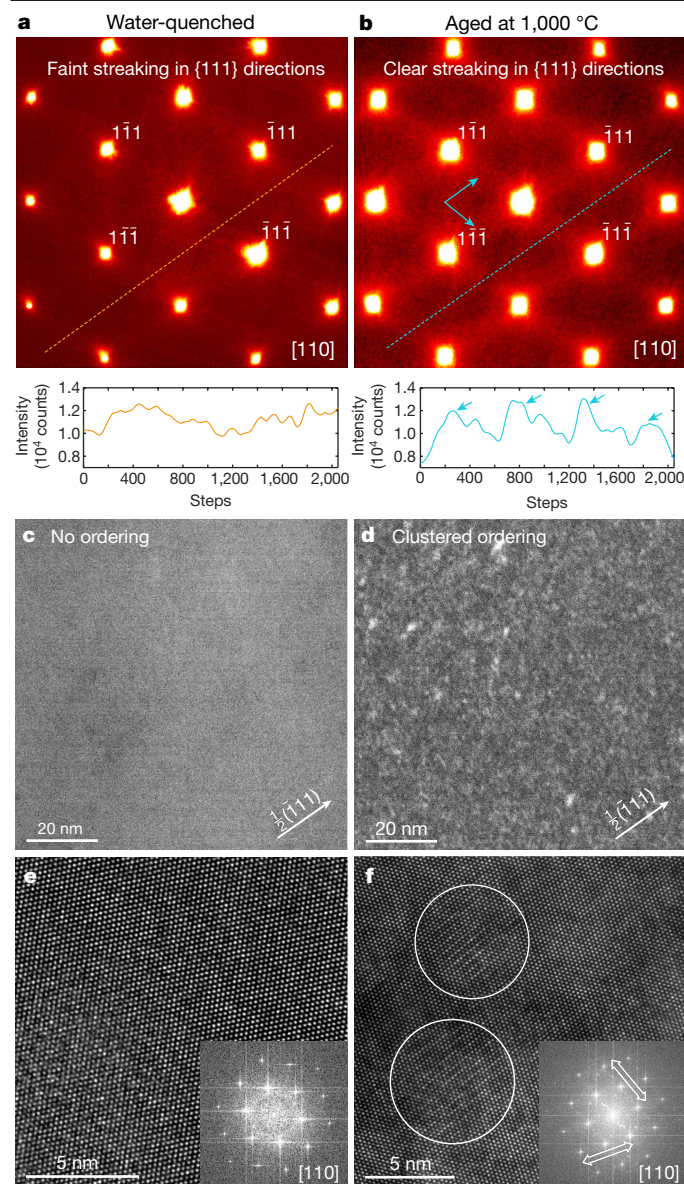


Fig. 1 | Energy-filtered TEM diffraction patterns, dark-field images formed with diffuse superlattice streaks and the associated high-resolution TEM images. Left column, water-quenched samples; right column, samples aged at 1,000 °C. **a, b**, Main panels, energy-filtered diffraction patterns for quenched and aged samples, respectively. The contrast is pseudo-coloured for better visibility. Plot below shows intensity measured along the diagonal line in the main panel; the periodic intensity of the diffuse superlattice streaks in **b** is marked by arrows. **c, d**, Energy-filtered dark-field images for quenched and aged samples, respectively. The aperture positions are marked by the **g** vectors (white arrows). **e, f**, Typical high-resolution TEM images of quenched and aged samples, respectively. Inset in each image is the associated FFT image. The features suggesting a superlattice are marked by the white circles, and the associated streaking along the {111} directions is marked by the white arrows in the FFT image.

SRO domains directly. While no dark-field contrast can be seen from the water-quenched samples (Fig. 1c), the aged sample (Fig. 1d) clearly reveals nanoscale domains. Results from an intermediate heat treatment are shown in Extended Data Fig. 1 for comparison.

The diffuse scattering in the diffraction patterns and the associated contrast in the dark-field images could arise from a combination of effects, including static and thermal displacement scattering and chemical SRO²⁴. In the CrCoNi system, the very close values of atomic

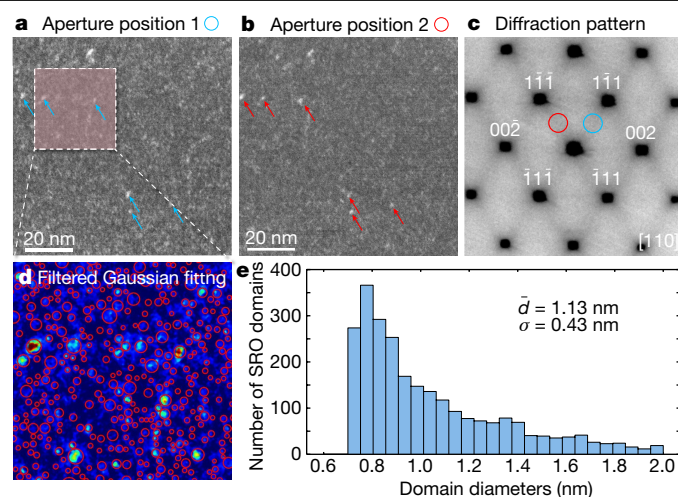


Fig. 2 | Evidence for the three-dimensional structure of the domains and their size distribution. **a, b**, Energy-filtered dark-field images from different diffuse superlattice peaks; examples showing the same domain contrast are marked with the arrows. **c**, Energy-filtered diffraction patterns of the region of interest; the red and blue circles indicate the dark-field imaging conditions of **a** and **b**. The contrast is reversed for better visibility. **d**, Magnified view of the boxed part of the dark-field image in **a**, with identified SRO domains marked by the red circles. The dark-field image is pseudo-coloured for better visibility. **e**, The histogram of identified domain diameters. The average value \bar{d} and the standard deviation σ are listed in the box.

scattering factors of the three elements would limit the contrast from any superlattice diffraction. However, the fact that the water-quenched samples (Fig. 1c) show negligible contrast using the same imaging conditions, and the fact that the aged samples show enhanced streaking and dark-field contrast, strongly support the interpretation that these features arise from the distortion of the local lattice associated with the formation of a diffuse SRO superlattice. Specifically, the enhanced contrast in samples aged at higher temperature and slow-cooled can be interpreted to be associated with the higher mobility of the atoms at these temperatures; this higher mobility enables the alloy to evolve towards a lower free-energy state with higher chemical SRO. Further evidence in support of this interpretation follows.

High-resolution TEM imaging (HRTEM) has been used to distinguish the difference between diffuse scattering induced by thermal displacement as compared to that induced by static displacement in previous studies²⁴. Figure 1e, f shows a comparison of HRTEM images from water-quenched and aged samples, where two regions in the aged sample show diffuse superlattice features along {111} planes as marked in Fig. 1f. In addition, the two-dimensional fast Fourier transforms (FFTs) of the HRTEM images (Fig. 1e, f insets) show a similar streaking intensity along the reciprocal lattice vectors that are normal to the {111} planes of the crystal. These observations provide clear evidence that the contrast in the real-space HRTEM images is associated directly with the diffuse intensity observed in the diffraction patterns. The features observed in the HRTEM images are qualitatively consistent with the type of order suggested in EXAFS²³ and in previous Monte Carlo simulations^{6,25}, both of which indicate that Cr–Cr pairs are strongly disfavoured at nearest-neighbour distances. Such bonding preferences are consistent with the alternating contrast caused by lattice distortion in the SRO domains along the $\langle 111 \rangle$ directions observed by HRTEM.

The combined conclusion from diffraction contrast and HRTEM imaging is that the high-temperature ageing leads to the formation of appreciable SRO in CrCoNi MEAs. The size and shape of the SRO-enhanced domains can thus be evaluated through energy-filtered dark-field imaging. For example, Fig. 2a, b presents two dark-field

images formed by using two different objective aperture positions as marked in Fig. 2c. While each dark-field image (Fig. 2a, b) shows mostly different sets of SRO-enhanced domains that are preferentially scattering to different parts of reciprocal space, there are a number of domains that could be identified in both images (examples are marked by the arrows). The existence of the same domains in images formed by separate and non-parallel directions of SRO-generated streaking is evidence for a non-planar shape of the SRO domains.

It is also possible to characterize the size distribution of the domains by assuming a shape (in this case we assume a spherical shape for simplicity) and applying a Gaussian template fitting algorithm²⁶ as demonstrated in the Methods section. This analysis generates an average diameter of the measured domains of 1.13 ± 0.43 nm, which would correspond to the third to fourth atomic shells on the fcc lattice of CrCoNi MEA^{17,20,27}. However, as the dark-field images in Figs. 1 and 2 suggest, the domain boundaries are relatively diffuse, and there is no evidence of any specific shape that characterizes the SRO domains. Further evidence for the diffuse nature of the SRO domains can be obtained by conducting geometrical phase analysis (GPA) on drift-corrected high-resolution scanning transmission electron microscope (STEM) images²⁸. The resulting strain maps are summarized in Extended Data Fig. 2. In the water-quenched sample, the fluctuation of local strain is minimal. However, in the sample aged at 1,000 °C, domain contrast similar in size to that found in the dark-field images could be identified, indicating small yet locally ordered fluctuations in lattice distortions. The results suggest that the SRO may be associated with the changes in the static atomic displacements, which is of interest since lattice distortions are widely proposed to partially explain the mechanical properties of the CrCoNi MEA¹³. This result thus warrants further investigation. We note, however, that standard X-ray diffraction (XRD) experiments conducted on both water-quenched and 1,000 °C aged samples show no evident changes in peak broadening for the two different thermal treatments (Extended Data Fig. 3), such that further investigations of the lattice distortions would probably require synchrotron measurements and lie beyond the scope of the present study.

It is known that the formation of SRO has a strong impact on dislocation plasticity—an increasing degree of SRO tends to increase the planarity of dislocation slip^{29–31}. To assess the effect in the CrCoNi alloy, dislocation analysis was conducted on bulk compressed samples and the results are summarized in Fig. 3. Specifically, a random distribution of dislocations was observed in the water-quenched sample, whereas a marked trend of localized planar configuration of dislocations was present in the 1,000 °C aged sample with SRO (Fig. 3a, b). In the latter case, the leading dislocations also tend to form dislocation pairs, where the separation distance of two adjacent dislocations was reduced (two examples are marked by the white arrows in Fig. 3b). One possible origin of planar slip in fcc materials is the Shockley partial dissociation of perfect dislocation cores, limiting the ability of dislocations to cross slip owing to the expanded cores. In the current study, however, the aged alloy possesses dislocation cores that are more compact than the quenched alloy while presenting planar slip. On the other hand, localized planar slip and leading dislocation pairs are usually correlated to the glide plane softening effect due to the local destruction of the SRO structure^{29,31,32}, where the initial dislocation motion interrupts the SRO atomic configuration and overcomes the energy barrier associated with the creation of a diffuse-anti-phase boundary (DAPB). Subsequently, dislocations following the initial dislocation would experience a lower energy barrier by gliding on the same path and avoiding the DAPB energy barrier. The DAPB energy as a function of dislocation slip events has been assessed by density functional theory (DFT) calculations based on the calculated SRO atomic configuration⁶, supporting this theory of the origin of the planar dislocation slip (Extended Data Fig. 4).

The exceptional strength, ductility and toughness of CrCoNi MEA can be directly correlated with the SFE of the material¹³. Previous DFT-assisted Monte Carlo simulations predicted that the SFE of CrCoNi

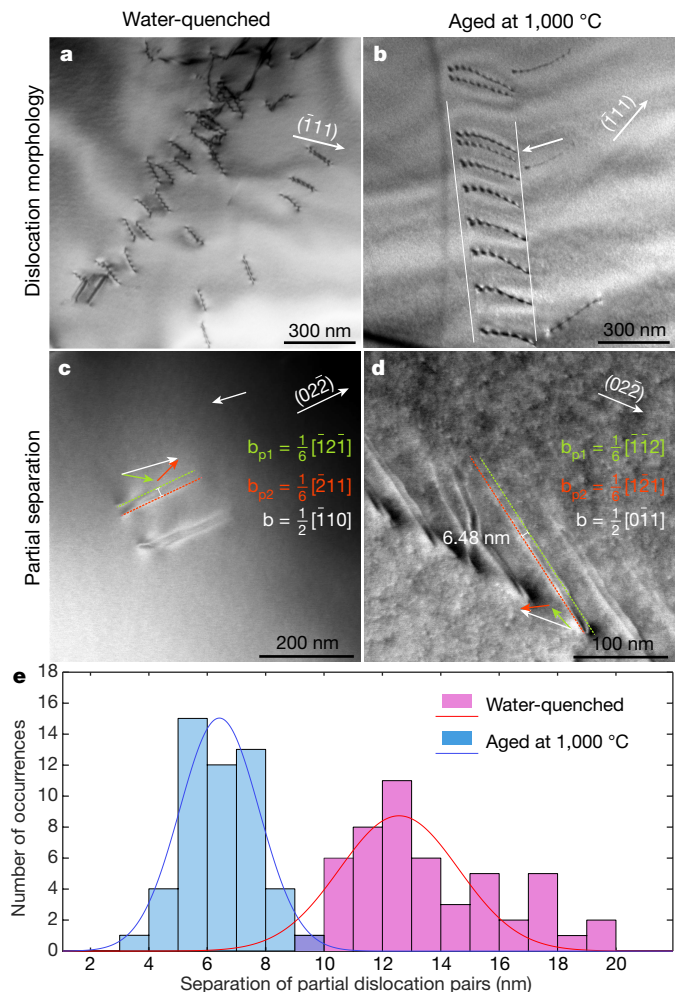


Fig. 3 | Dislocation analysis of both water-quenched and 1,000 °C aged samples. **a**, Two-beam bright-field image, showing the representative wavy configuration of dislocations in the water-quenched sample. The white arrow with the **g** vector marks the two-beam diffraction condition utilized. **b**, Two-beam bright-field image, showing the representative planar configuration (marked by the parallel white lines) of dislocations in the 1,000 °C aged sample. The leading dislocation pair is marked by the white arrow. The white arrow with the **g** vector marks the two-beam condition utilized. **c, d**, Low angle annular dark field (LAADF) images showing dislocation dissociations in water-quenched and 1,000 °C aged samples, respectively. The white arrows with the **g** vectors mark the two-beam diffraction conditions utilized. The Burgers vector relations are demonstrated: green arrows, \mathbf{b}_{p1} ; red arrows, \mathbf{b}_{p2} ; white arrows, \mathbf{b} . The detailed '**g**·**b**' analysis is summarized in Extended Data Fig. 5. Examples of measured partial dislocation separations are marked as 12.74 nm in **c** and 6.48 nm in **d**. **e**, Distribution of the measured separation of partial dislocation pairs from both water-quenched and 1,000 °C aged samples. The results of numerical analysis are summarized in Extended Data Table 1.

MEA could be highly tunable by varying the SRO⁶. While the SFE of MEA/HEAs has been experimentally probed previously via both weak-beam dark-field imaging²⁰ and diffraction contrast STEM (DC-STEM) analysis²², the SFE has never been directly correlated with the degree of SRO. In the current study, the SFE was measured by DC-STEM analysis as the technique allows imaging through thicker samples to minimize the sample surface effect. Figure 3c, d shows examples of images where partial dislocations could be identified and their disassociation measured directly (analysis detailed in Extended Data Fig. 5). The separation distance and the statistical results are summarized in Fig. 3e and Extended Data Table 1. The detailed calculation of the SFE is elaborated

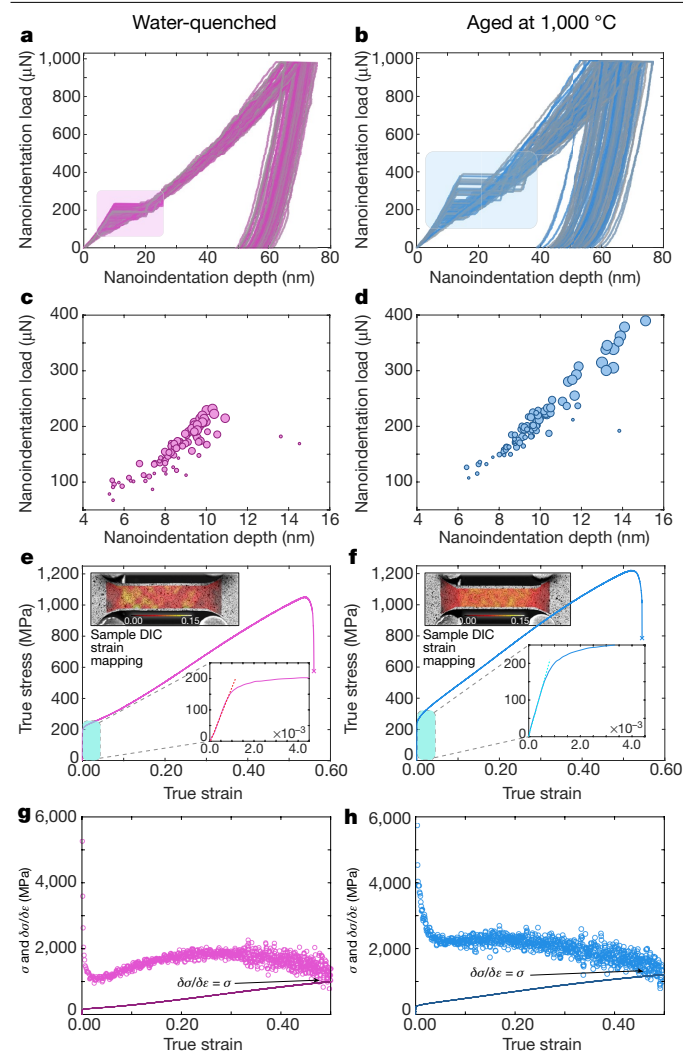


Fig. 4 | Comparison of mechanical properties from nanoindentation and bulk tensile tests. Left column, water-quenched samples; right column, samples aged at 1,000 °C. **a, b**, Load–depth curves from a 10×10 grid of nanoindentations separated by $10 \mu\text{m}$ from each other, from a water-quenched sample and a 1,000 °C aged sample, respectively. Pop-in analysis from these same tests are provided for water-quenched (**c**) and 1,000 °C aged (**d**) samples. Data points (circles) depict the depth and load when the pop-in events occur; the sizes of the circles are proportional to the total pop-in displacement. **e, f**, Results of tensile tests on a water-quenched sample and a 1,000 °C aged sample, respectively. Lower insets, the elastic portions of the curves (note that the inset x axis shows true strain $\times 10^{-3}$); upper insets, a sample image of the strain distribution during elastic loading, as determined by digital image correlation (DIC). **g, h**, Work hardening rate derived from the true stress–strain curves (σ , true stress; ϵ , true strain; $d\sigma/d\epsilon$, calculated work hardening rate) of the water-quenched and the 1,000 °C aged samples, respectively. True stress versus true strain data from the same tests, respectively, are also displayed in **g, h** as the solid lines for comparison. The necking points ($d\sigma/d\epsilon = \sigma$) are marked by the black arrows. The results of numerical analysis from these tests are summarized in Extended Data Table 1.

in the Methods section, and shows that the 1,000 °C aged samples have an SFE of $23.33 \pm 4.31 \text{ mJ m}^{-2}$, double the value of its water-quenched counterpart ($8.18 \pm 1.43 \text{ mJ m}^{-2}$). This measurement confirms that the SRO directly impacts the SFE, and indicates that the SFE could be fine-tuned by controlling the ordering⁶.

In order to quantify the impact of SRO on the mechanical properties of the CrCoNi MEA, both nanoindentation tests and bulk tensile tests were performed. The measured nanoindentation hardness is

$4.07 \pm 0.23 \text{ GPa}$ for the water-quenched sample and $4.37 \pm 0.58 \text{ GPa}$ for the 1,000 °C aged sample. SRO also significantly affects the onset of plasticity, which is manifested by the ‘pop-in’ event³³ in the load versus displacement curves in Fig. 4c, d. The first pop-in events of the 1,000 °C aged sample are distributed more discretely and usually occur at higher load than the quenched sample. In addition, the displacement plateau that corresponds to the strain burst of a pop-in event is larger in the aged material, as detailed in Extended Data Fig. 6. The higher pop-in load and larger displacement plateau in the 1,000 °C aged specimen indicates the presence of dislocation avalanches (sudden bursts of dislocation nucleation and propagation), providing further evidence of the SRO hardening and the subsequent glide plane softening caused by passage of the first few dislocations in the slip band. Bulk tensile tests confirmed the strengthening effect of SRO by showing an approximately 25% increase of the yield strength (Extended Data Table 1) as well as a marked change of the work hardening behaviour.

As demonstrated in Fig. 4g, h, the initial work hardening rate of the aged sample is twice that of its water-quenched counterpart, reinforcing the that the hardening is caused by the SRO domains. Traditionally, the formation of SRO in alloys causes planar dislocation slip and deformation localization^{29,34–36}. In some cases, the deformation localization affects the alloys’ ductility and toughness, whereas in the current study, the formation of SRO has little effect on the overall ductility of the MEA alloy. Deformation twinning is reported to explain the exceptional ductility of the CrCoNi alloy^{13,15}, in which nano-twinning delays deformation localization. Though direct evidence is lacking, when we consider the similar work-hardening behaviour at later stages of deformation of both the 1,000 °C aged and the water-quenched samples, we speculate that the exceptional strength and toughness of CrCoNi MEA arises in part from this unique combination of SRO hardening and twin-induced deformation at later stages. However, further systematic analysis is required to fully understand any potential effect of SRO on the deformation twinning.

As an emerging class of structural materials, MEA/HEAs possess a desirable combination of mechanical properties for structural applications^{13,37,38}. Although the concept of MEA/HEAs is based on production of a single-phase solid solution, there has long been a question about how well-mixed the solid solutions are^{4,8,13,23,39–43}. Here we directly imaged the local ordering and showed how the deformation behaviour of MEAs is directly correlated with the degree of SRO. Annealing the MEA to promote SRO led to an increase in hardness, a doubling of the SFE and a subsequent increase in planar slip. Owing to its impact on the mechanical properties, the degree of SRO is a critical feature that should be considered in the materials’ design phase. Directly tailoring the SRO microstructure on an atomic level therefore provides another route for controlling the structure–property relationship of advanced materials.

Online content

Any methods, additional references, Nature Research reporting summaries, source data, extended data, supplementary information, acknowledgements, peer review information; details of author contributions and competing interests; and statements of data and code availability are available at <https://doi.org/10.1038/s41586-020-2275-z>.

1. Yeh, J. W. et al. Nanostructured high-entropy alloys with multiple principal elements: novel alloy design concepts and outcomes. *Adv. Eng. Mater.* **6**, 299–303 (2004).
2. Cantor, B., Chang, I. T. H., Knight, P. & Vincent, A. J. B. Microstructural development in equiatomic multicomponent alloys. *Mater. Sci. Eng. A* **375–377**, 213–218 (2004).
3. Li, Z., Pradeep, K. G., Deng, Y., Raabe, D. & Tasan, C. C. Metastable high-entropy dual-phase alloys overcome the strength–ductility trade-off. *Nature* **534**, 227–230 (2016).
4. Gludovatz, B. et al. A fracture-resistant high-entropy alloy for cryogenic applications. *Science* **345**, 1153–1158 (2014).
5. Yang, T. et al. Multicomponent intermetallic nanoparticles and superb mechanical behaviors of complex alloys. *Science* **362**, 933–937 (2018).
6. Ding, J., Yu, Q., Asta, M. & Ritchie, R. O. Tunable stacking fault energies by tailoring local chemical order in CrCoNi medium-entropy alloys. *Proc. Natl Acad. Sci. USA* **115**, 8919–8924 (2018).

7. Li, Q.-J., Sheng, H. & Ma, E. Strengthening in multi-principal element alloys with local-chemical-order roughened dislocation pathways. *Nat. Commun.* **10**, 3563 (2019).
8. Zhang, Y. et al. Microstructures and properties of high-entropy alloys. *Prog. Mater. Sci.* **61**, 1–93 (2014).
9. Gao, M. C., Yeh, J. W., Liaw, P. K. & Zhang, Y. *High-Entropy Alloys: Fundamentals and Applications* (Springer International, 2016).
10. Senkov, O. N., Miracle, D. B., Chaput, K. J. & Couzinie, J. P. Development and exploration of refractory high entropy alloys — a review. *J. Mater. Res.* **33**, 3092–3128 (2018).
11. Miracle, D. B. High entropy alloys as a bold step forward in alloy development. *Nat. Commun.* **10**, 1805 (2019).
12. George, E. P., Raabe, D. & Ritchie, R. O. High-entropy alloys. *Nat. Rev. Mater.* **4**, 515–534 (2019).
13. Li, Z., Zhao, S., Ritchie, R. O. & Meyers, M. A. Mechanical properties of high-entropy alloys with emphasis on face-centered cubic alloys. *Prog. Mater. Sci.* **102**, 296–345 (2019).
14. Gludovatz, B. et al. Exceptional damage-tolerance of a medium-entropy alloy CrCoNi at cryogenic temperatures. *Nat. Commun.* **7**, 10602 (2016).
15. Zhang, Z. et al. Dislocation mechanisms and 3D twin architectures generate exceptional strength-ductility-toughness combination in CrCoNi medium-entropy alloy. *Nat. Commun.* **8**, 14390 (2017).
16. Zaddach, A. J., Niu, C., Koch, C. C. & Irving, D. L. Mechanical properties and stacking fault energies of NiFeCrCoMn high-entropy alloy. *J. Miner. Met. Mater. Soc.* **65**, 1780–1789 (2013).
17. Zhang, Y. H., Zhuang, Y., Hu, A., Kai, J. J. & Liu, C. T. The origin of negative stacking fault energies and nano-twin formation in face-centered cubic high entropy alloys. *Scr. Mater.* **130**, 96–99 (2017).
18. Zhao, S., Stocks, G. M. & Zhang, Y. Stacking fault energies of face-centered cubic concentrated solid solution alloys. *Acta Mater.* **134**, 334–345 (2017).
19. Niu, C., LaRosa, C. R., Miao, J., Mills, M. J. & Ghazisaeidi, M. Magnetically-driven phase transformation strengthening in high entropy alloys. *Nat. Commun.* **9**, 1363 (2018).
20. Laplanche, G. et al. Reasons for the superior mechanical properties of medium-entropy CrCoNi compared to high-entropy CrMnFeCoNi. *Acta Mater.* **128**, 292–303 (2017).
21. Okamoto, N. L. et al. Size effect, critical resolved shear stress, stacking fault energy, and solid solution strengthening in the CrMnFeCoNi high-entropy alloy. *Sci. Rep.* **6**, 35863 (2016).
22. Smith, T. M. et al. Atomic-scale characterization and modeling of 60° dislocations in a high-entropy alloy. *Acta Mater.* **110**, 352–363 (2016).
23. Zhang, F. X. et al. Local structure and short-range order in a NiCoCr solid solution alloy. *Phys. Rev. Lett.* **118**, 205501 (2017).
24. Van Tendeloo, G. & Amelinckx, S. The origin of diffuse intensity in electron diffraction patterns. *Phase Transit.* **67**, 101–135 (1998).
25. Tamm, A., Aabloo, A., Klintonberg, M., Stocks, M. & Caro, A. Atomic-scale properties of Ni-based FCC ternary, and quaternary alloys. *Acta Mater.* **99**, 307–312 (2015).
26. Zhang, R. et al. Direct imaging of short-range order and its relationship to deformation in Ti-6Al. *Sci. Adv.* **5**, eaax2799 (2019).
27. Laplanche, G. et al. Elastic moduli and thermal expansion coefficients of medium-entropy subsystems of the CrMnFeCoNi high-entropy alloy. *J. Alloys Compd.* **746**, 244–255 (2018).
28. Ophus, C., Ciston, J. & Nelson, C. T. Correcting nonlinear drift distortion of scanning probe and scanning transmission electron microscopies from image pairs with orthogonal scan directions. *Ultramicroscopy* **162**, 1–9 (2016).
29. Gerold, V. & Karthaler, H. P. On the origin of planar slip in f.c.c. alloys. *Acta Metall.* **37**, 2177–2183 (1989).
30. Pekin, T. C., Gammer, C., Ciston, J., Ophus, C. & Minor, A. M. In situ nanobeam electron diffraction strain mapping of planar slip in stainless steel. *Scr. Mater.* **146**, 87–90 (2018).
31. Neeraj, T. & Mills, M. J. Short-range order (SRO) and its effect on the primary creep behavior of a Ti-6wt.%Al alloy. *Mater. Sci. Eng. A* **319–321**, 415–419 (2001).
32. van de Walle, A. & Asta, M. First-principles investigation of perfect and diffuse antiphase boundaries in HCP-based Ti-Al alloys. *Metall. Mater. Trans. A* **33**, 735–741 (2002).
33. Gouldstone, A. et al. Indentation across size scales and disciplines: recent developments in experimentation and modeling. *Acta Mater.* **55**, 4015–4039 (2007).
34. Fisher, J. C. C. On the strength of solid solution alloys. *Acta Metall.* **2**, 9–10 (1954).
35. Britton, T. B., Dunne, F. P. E. & Wilkinson, A. J. On the mechanistic basis of deformation at the microscale in hexagonal close-packed metals. *Proc. R. Soc. A* **471**, 20140881 (2015).
36. Hamdi, F. & Asgari, S. Influence of stacking fault energy and short-range ordering on dynamic recovery and work hardening behavior of copper alloys. *Scr. Mater.* **62**, 693–696 (2010).
37. Miracle, D. B. et al. Exploration and development of high entropy alloys for structural applications. *Entropy* **16**, 494–525 (2014).
38. Williams, J. C. & Starke, E. A. Progress in structural materials for aerospace systems. *Acta Mater.* **51**, 5775–5799 (2003).
39. Ma, Y. et al. Chemical short-range orders and the induced structural transition in high-entropy alloys. *Scr. Mater.* **144**, 64–68 (2018).
40. Singh, P., Smirnov, A. V. & Johnson, D. D. Atomic short-range order and incipient long-range order in high-entropy alloys. *Phys. Rev. B* **91**, 224204 (2015).
41. Lucas, M. S. et al. Absence of long-range chemical ordering in equimolar FeCoCrNi. *Appl. Phys. Lett.* **100**, 251907 (2012).
42. Niu, C. et al. Spin-driven ordering of Cr in the equiatomic high entropy alloy NiFeCrCo. *Appl. Phys. Lett.* **106**, 161906 (2015).
43. Linden, Y., Pinkas, M., Munitz, A. & Meshi, L. Long-period antiphase domains and short-range order in a B2 matrix of the AlCoCrFeNi high-entropy alloy. *Scr. Mater.* **139**, 49–52 (2017).

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

© This is a U.S. government work and not under copyright protection in the U.S.; foreign copyright protection may apply 2020

Methods

Materials and sample preparation

The raw ingot of the equiatomic CrCoNi MEA was argon-arc double melted and then cut into smaller samples. The samples were then divided into two groups and underwent different thermal treatments: (1) homogenized at 1,200 °C for 48 h then water quenched to room temperature (uniform texture, grain size ~800 µm as determined by electron backscatter diffraction, EBSD); or (2) homogenized at 1,200 °C for 48 h then aged at 1,000 °C for 120 h followed by furnace cooling (uniform texture, grain size ~1,000 µm as determined by EBSD). All the alloys were confirmed to be a single-phase fcc structure via X-ray diffraction and EBSD analysis. Samples for dislocation analysis were further deformed by conducting bulk compression tests on an MTS Criterion (Model 43) system to introduce dislocation plasticity. The final strain was 6% with a strain rate of 1×10^{-3} . The samples were then sliced and thinned by mechanical polishing. Electron-transparent samples for TEM observation were prepared with a Fischione twin-jet electropolisher using a solution of 70% methanol, 20% glycerol and 10% perchloric acid at -20 °C. The samples for nanoindentation tests were prepared by single-side electrochemical polishing with the aforementioned solution and parameters.

Energy-filtered TEM and SRO recognition

TEM samples of different heat treatments were used for observation. A Zeiss LIBRA 200MC microscope, equipped with an in-column Ω energy filter, was used to take both diffraction patterns and dark-field images. It is necessary to consider the impact of the objective aperture on the resolution, which could be estimated by the Airy disk radius using the following equation⁴⁴:

$$r_{\text{Airy}} \approx \frac{1.2\lambda f}{D} \quad (1)$$

where λ is the electron wavelength (0.02507 Å for 200 kV TEM), f is the focal length of the objective lens (~3 mm for the Zeiss LIBRA) and D is the diameter of the objective aperture (25 µm in the current study). For the experimental setup used in the current study, the size of the aperture Airy disk is 3.61 Å, which is below the size of the observed SRO domains. An alternative way to determine the resolution limit is to directly measure the semi-angle of the aperture used:

$$r_{\text{Airy}} \approx \frac{1.2\lambda}{\alpha} = 1.2d' \quad (2)$$

where α is the measured semi-angle of the aperture and d' is the measured size of the aperture in reciprocal space. This method yields a similar resolution limit of 3.03 Å, confirming sufficient resolution to resolve the SRO domains. A 5-eV energy slit was deployed to select the zero-loss peak and eliminate the contrast from inelastic scattering. A Gatan US1000 CCD camera was used to acquire the diffraction patterns and dark-field images. Before the data analysis, the energy-filtered dark-field images were filtered by a dark reference subtraction. According to the energy-filtered dark-field image shown in Figs. 1 and 2, there is no observable directional tendency of the domains. Therefore, we assumed a circular kernel signal from the domains for our analysis. SRO-enhanced domains were identified and measured through Gaussian template fitting, where 2D convolutions with the dark-field image were conducted using a list of differently sized 2D Gaussian templates (with different values of standard deviation²⁶). The stack of result images was further analysed through a circular Hough transform to identify all signal peaks. The intensity cutoff was set according to the best fit result. Overlapping entities were deleted to ensure an accurate size measurement. Details of the algorithm are given in Extended Data Table 2.

A manual sampling was carried out to estimate the domain sizes and gain a reference for the optimization of parameters. Two critical

parameters that would impact the identification are the minimal signal cutoff and the domain diameter range. The optimization process was conducted according to the best fitting results. In the case of a high signal cutoff or a narrow diameter range, the algorithm will miss some of the major contrast, whereas, in the case of a low signal cutoff or a wide diameter range, the algorithm will pick up many small intensity fluctuations that are from camera noise. It is worth mentioning the limitations to the domain recognition algorithm. Specifically, the assumption that the domains are spherical is for simplification, but the shapes of the domains vary. Parallel attempts using a threshold segmentation algorithm involved much more subjectivity and yielded unreasonable results. However, the purpose of the analysis is to provide an estimated size distribution of the SRO domains, for which the current analysis is sufficient until large-scale atomic imaging studies can provide similar statistics.

X-ray diffraction (XRD) experiments

The XRD experiments were performed ex situ with a PANalytical XPert diffractometer on water-quenched and on 1,000 °C aged samples. The scan range (2θ) was set to 42°–54° to include the (111) and the (200) peaks. The angle resolution was set to 0.005° with a 0.8-s integration time to ensure an accurate measurement of the lattice constants.

High-resolution STEM (HRSTEM) imaging

HRSTEM imaging of water-quenched and 1,000 °C aged samples were conducted on the double-corrected TEAM I microscope (operated at 300 kV) at the National Center for Electron Microscopy (NCEM), Lawrence Berkeley National Laboratory. Drift correction was conducted with the methods developed by Ophus et al.²⁸ to eliminate artefacts from beam scan jittering. The FRWRtools plugin for Gatan Digital Micrograph software were used for the following GPA analysis. Averaged fast-Fourier transforms were used as strain templates. The real-space resolution was set to 1.5 nm to achieve a relatively accurate measurement in reciprocal space.

STEM EDS measurements

Quantitative energy dispersive X-ray mapping (energy dispersive X-ray spectroscopy, EDS) was conducted on both the water-quenched samples and aged samples using a TitanX microscope with a quad EDS detector. No chemical segregation was observed; results are summarized in Extended Data Fig. 7. The lack of any visible chemical segregation via EDS analysis in the aged samples is consistent with the HRSTEM observation presented in Extended Data Fig. 2, where there is no obvious Z-contrast difference despite different degrees of local lattice distortion. Previous theoretical studies^{6,7} revealed that the SRO in the CrCoNi MEA is in the range of several nearest neighbour distances and that the driving force for the formation of SRO is to avoid certain types of bonding. Combined with the observation presented in the current study, we can conclude that it is not necessary for the SRO structure to possess a strong chemical segregation. Further verification using atomic-resolution EDS or electron energy loss spectroscopy (EELS) could provide valuable insights revealing the atomic structure of SRO clusters.

Dislocation analysis

TEM dislocation analysis was conducted on both the water-quenched and the aged samples after 6% compressive deformation. TEM observations were conducted on the Zeiss LIBRA 200MC (operating at 200 kV) at NCEM. Low-angle annular dark-field DC-STEM images^{45,46} for ' $\mathbf{g}\cdot\mathbf{b}$ ' analysis of the partial dislocations and SFE measurements were acquired on the TEAM I microscope. To identify the Burgers vectors of the partial dislocations, ' $\mathbf{g}\cdot\mathbf{b}$ ' analysis was performed where the contrast from a dislocation is eliminated (or minimized) by using a diffraction condition normal to the Burgers vector such that ' $\mathbf{g}\cdot\mathbf{b} = 0$ '. The measured partial dislocation separation was further calibrated by conducting a $g(3g)$ weak-beam

dark-field imaging and calculating the actual partial separation from the observed values^{20,47,48}. The SFEs were calculated according to the following equation^{20,49,50}:

$$\text{SFE} = \frac{Gb_p^2}{8\pi d} \left(\frac{2-\nu}{1-\nu} \right) \left(1 - \frac{2\nu \cos(2\beta)}{2-\nu} \right) \quad (3)$$

where G is the shear modulus of the CrCoNi MEA (determined by the ultrasonic pulse-echo measurement), b_p is the magnitude of the Burgers vector of partial dislocations (~ 0.146 nm), d is the measured separation of partial dislocations, ν is Poisson's ratio (determined by the ultrasonic pulse-echo measurement), and β is the angle between the perfect dislocation Burgers vector and the dislocation line. For both the 1,000 °C aged samples and water-quenched samples, 50 individual measurements were conducted on more than 10 partial pairs from relatively thick regions to avoid any surface effects. Associated \pm standard deviations were calculated to ensure accurate and representative results.

Nanoindentation experiments

Nanoindentation tests were conducted on a Bruker Ti 950 TriboIndenter instrument with a 1- μ m Berkovich tip. The peak load was set to 1,000 μ N. The analysis was conducted with a calibrated area function of the tip. The water-quenched and 1,000 °C aged samples were electrochemically polished on one side with a solution of 70% methanol, 20% glycerol and 10% perchloric acid at -20 °C. A 10×10 grid of indentations covering an area of 1 mm \times 1 mm was set to conduct the test for each sample. No strong texture was observed by post-test EBSD. All quantitative parameters were averaged over the 100 indentations with associated \pm standard deviations.

Bulk mechanical tests

Bulk tensile tests were carried out on an MTS Criterion (Model 43) system. A Sony A7R Mark II camera was used to record images for Digital Image Correlation (DIC). A copy of Vic-2D Image Correlation software was used to conduct the DIC analysis. Owing to the limited amount of material, the dimension of the gauge section of both water-quenched and 1,000 °C aged samples was set to 5.1 mm \times 0.8 mm \times 1.6 mm. Specially designed sample grippers were used to conduct the tensile test. Sample surfaces were mechanically polished and sparkle-sprayed before the tests. The strain was extracted from the DIC von Mises strain data using the 'virtual extensometers' mode and averaged three virtual extensometers along the gauge length.

Diffuse anti-phase boundary energy

The diffuse anti-phase boundary (DAPB) energy as a function of dislocation slip events was calculated via density functional theory using an 'aged' atomic model reported in previous literature⁶, which has an SFE similar to that of the 1,000 °C aged samples. Excess energy was calculated after each successive slip was introduced into the system.

Elastic modulus measurements

In addition to the effect of SRO on plastic behaviour, it also, in theory, should affect elastic properties as the local bonding environments are substantially altered from the perfect random solid solution. A simple rule-of-mixtures would predict a Young's modulus of ~ 229 GPa for equiatomic CrCoNi (ref. ³⁸). However, the nanoindentation modulus (reduced modulus) of the water-quenched sample is measured to be 181.76 ± 13.37 GPa, 18.1% smaller than that of the 1,000 °C aged sample (214.79 ± 18.49 GPa). In contrast, the global Young's modulus of the bulk materials was determined by the ultrasonic pulse-echo technique where the longitudinal and transverse sound speeds are measured to calculate elastic modulus. An Olympus 38DL Plus thickness gauge with a Model 5072PR pulser/receiver module was used to measure the

speed of the shear velocity and the longitudinal velocity. The Poisson's ratio, Young's modulus and the shear modulus were calculated with the following equations:

$$\nu = \frac{1 - 2(V_T/V_L)^2}{2 - 2(V_T/V_L)^2} \quad (4)$$

$$E = \frac{V_L^2 \rho (1 + \nu)(1 - 2\nu)}{1 - \nu} \quad (5)$$

$$G = V_L^2 \rho \quad (6)$$

where ν is the Poisson's ratio, V_T is the shear velocity, V_L is the longitudinal velocity, E is the Young's modulus, G is the shear modulus and ρ is the density of the materials, which is estimated with the following equation:

$$\rho = \frac{4m_{\text{aAvg}}}{V_{\text{cell}}} \quad (7)$$

where m_{aAvg} is the averaged atomic mass of Cr, Co and Ni, V_{cell} is the volume of an fcc unit cell calculated with the lattice constants derived from the XRD results.

The measured global Young's modulus of the water-quenched and the aged samples was 229.93 GPa and 230.99 GPa, respectively (other measured elastic properties are listed in Extended Data Table 1). The discrepancy between the locally-measured modulus by nanoindentation and the bulk-scale modulus measured acoustically may result from the limited size (~ 1 nm) of SRO clusters. The local measurement of modulus by nanoindentation is sensitive to the homogeneity of the distribution of the SRO clusters. However, the wavelength of the ultrasonic acoustic waves used to measure global modulus is orders of magnitude longer than the size of the SRO. Therefore, the measurement is averaged over a much larger volume and is insensitive to the degree of SRO.

Data availability

The data that support the findings of this study are available from the corresponding author upon reasonable request.

44. Fultz, B. & Howe, J. M. *Transmission Electron Microscopy and Diffractometry of Materials* (Springer, 2008).
45. Phillips, P. J., Brandes, M. C., Mills, M. J. & de Graef, M. Diffraction contrast STEM of dislocations: imaging and simulations. *Ultramicroscopy* **111**, 1483–1487 (2011).
46. Phillips, P. J. et al. Atomic-resolution defect contrast in low angle annular dark-field STEM. *Ultramicroscopy* **116**, 47–55 (2012).
47. Cockayne, D. J. H. The weak-beam technique as applied to dissociation measurements. *J. Phys. Colloq.* **35**, C7-141–C7-148 (1974).
48. Cockayne, D. J. H., Ray, I. L. F. & Whelan, M. J. Investigations of dislocation strain fields using weak beams. *Phil. Mag.* **20**, 1265–1270 (1969).
49. Anderson, P. M., Hirth, J. P. & Lothe, J. *Theory of Dislocations* (Cambridge Univ. Press, 2017).
50. Pierce, D. T. et al. The influence of manganese content on the stacking fault and austenite/ ϵ -martensite interfacial energies in Fe–Mn–(Al–Si) steels investigated by experiment and theory. *Acta Mater.* **68**, 238–253 (2014).

Acknowledgements This work was primarily supported by the US Department of Energy, Office of Science, Office of Basic Energy Sciences, Materials Sciences and Engineering Division, under contract no. DE-AC02-05-CH11231 within the Damage-Tolerance in Structural Materials (KC 13) programme. R.Z., S.Z. and Y.C. acknowledge support from the US Office of Naval Research under grant nos. N00014-12-1-0413, N00014-17-1-2283 and N00014-11-1-0886, respectively. Work at the Molecular Foundry was supported by the Office of Science, Office of Basic Energy Sciences, of the US Department of Energy under contract no. DE-AC02-05CH11231. X-ray diffraction measurements were made at the Stanford Nano Shared Facilities (SNSF), supported by the National Science Foundation under award ECCS-1542152. This research used resources of the National Energy Research Scientific Computing Center (NERSC), a US Department of Energy Office of Science User Facility operated under contract no. DE-AC02-05CH11231. We thank E. Ma at Johns Hopkins University for providing a 600 °C aged alloy.

Article

Author contributions R.Z., S.Z., M.A., R.O.R. and A.M.M. conceived the project.; R.Z. and S.Z. conducted the energy-filtered TEM imaging and dislocation analysis; C.O. and R.Z. developed and optimized the domain recognition algorithm; R.Z. and S.Z. conducted the nanoindentation tests; R.Z., S.Z and Y.C. conducted the tensile tests. J.D. conducted the DFT simulations. T.J. conducted the XRD experiments. R.Z., S.Z., R.O.R., M.A. and A.M.M. prepared the manuscript, which was reviewed and edited by all authors. Project administration, supervision and funding acquisition was performed by R.O.R., M.A. and A.M.M.

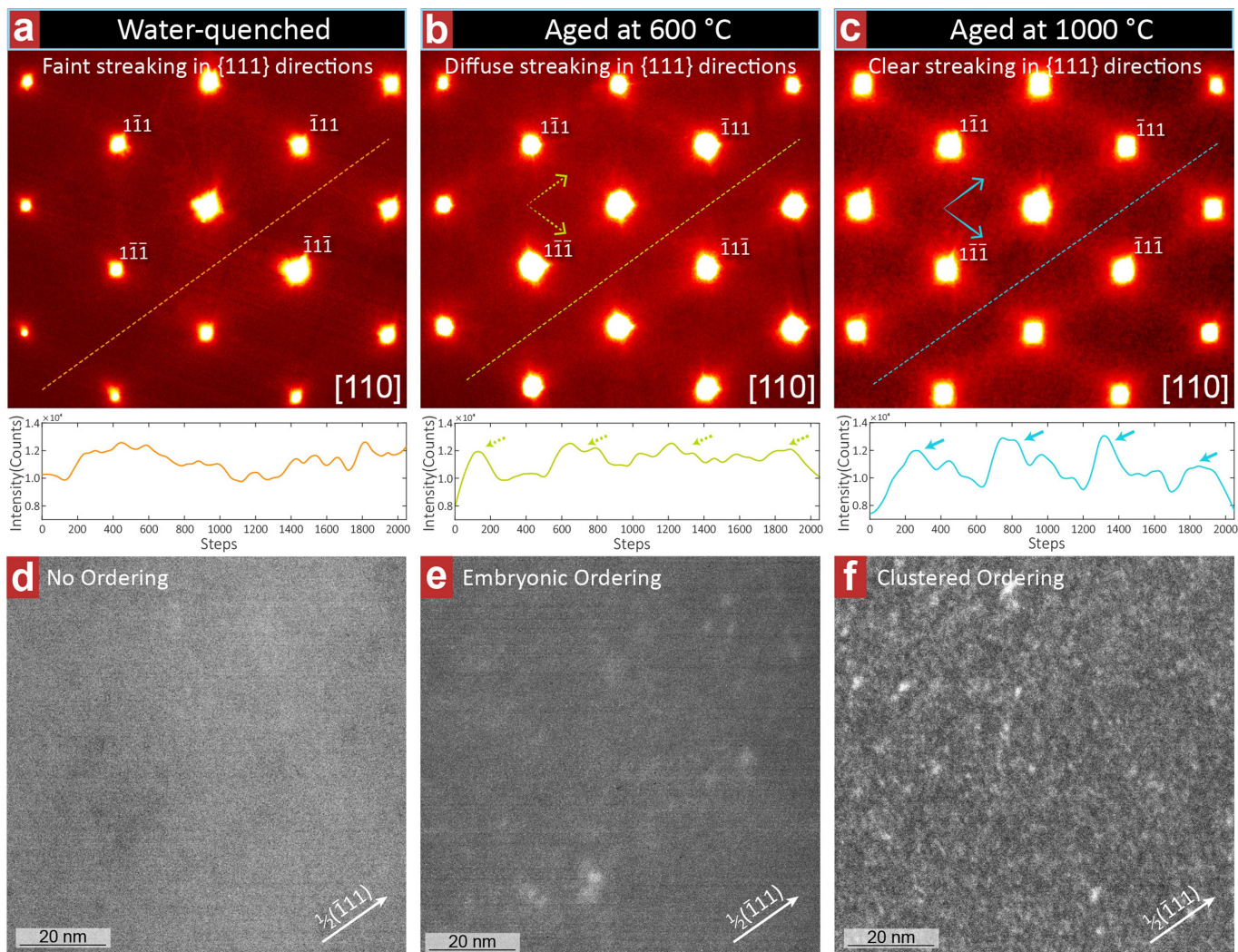
Competing interests The authors declare no competing interests.

Additional information

Correspondence and requests for materials should be addressed to A.M.M.

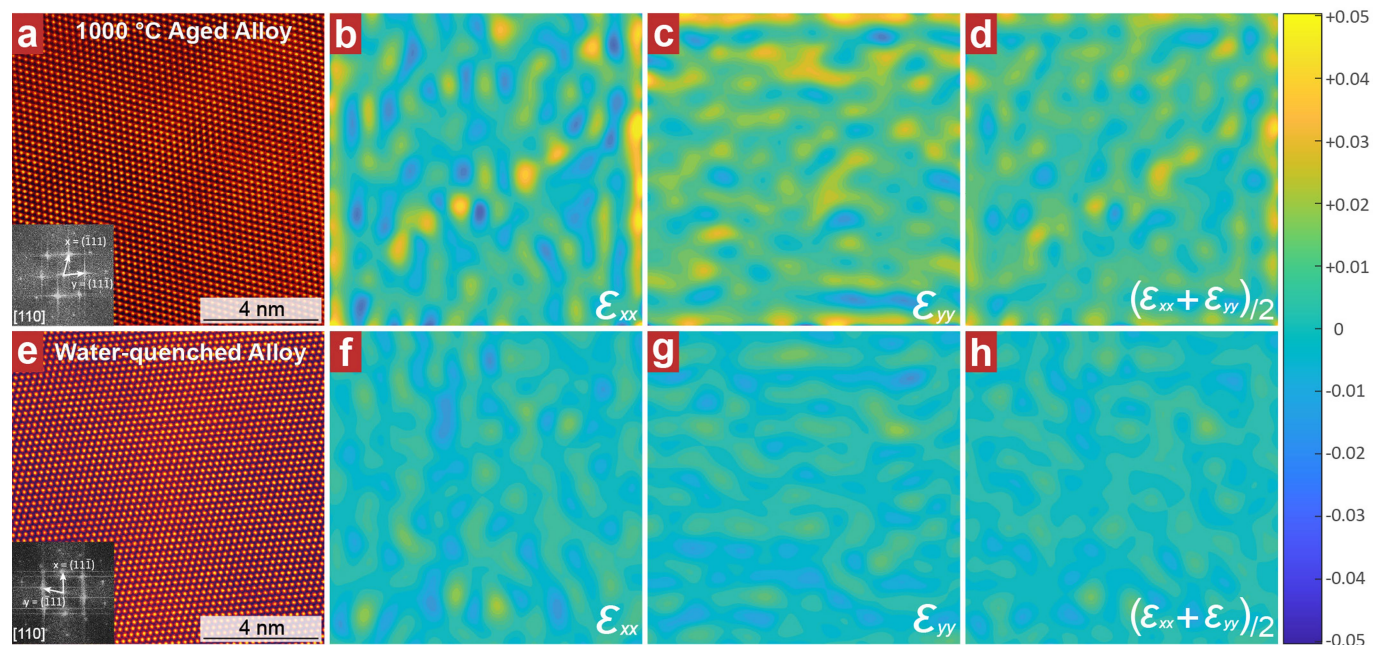
Peer review information *Nature* thanks Elena Pereloma, Christopher Woodward and the other, anonymous, reviewer(s) for their contribution to the peer review of this work.

Reprints and permissions information is available at <http://www.nature.com/reprints>.



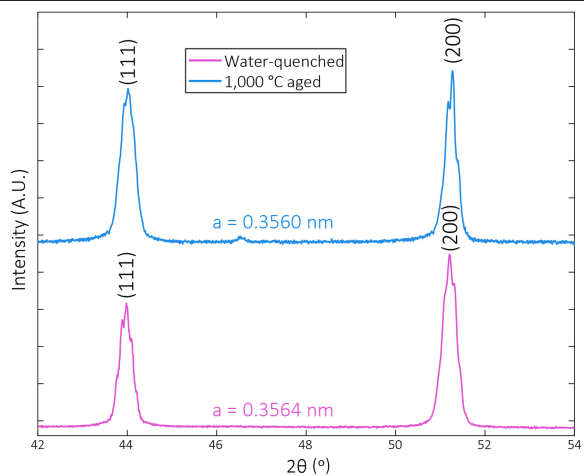
Extended Data Fig. 1 | Energy-filtered TEM diffraction patterns and dark-field images formed with diffuse superlattice streaks.
a–c, Energy-filtered diffraction patterns taken from CrCoNi MEA samples that were water-quenched, aged at 600 °C for one week or aged at 1,000 °C for one week, respectively. The contrast is pseudo-coloured for better visibility. The line plots of intensity show the periodic intensity of the diffuse superlattice

streaks. **d–f**, Energy-filtered dark-field images taken from water-quenched, 600 °C aged and 1,000 °C aged samples, respectively. The aperture positions are marked by the \mathbf{g} vectors (labelled arrows). The images of the water-quenched and the 1,000 °C aged samples are the same as in Fig. 1 but are presented again here for comparison with the 600 °C aged sample.

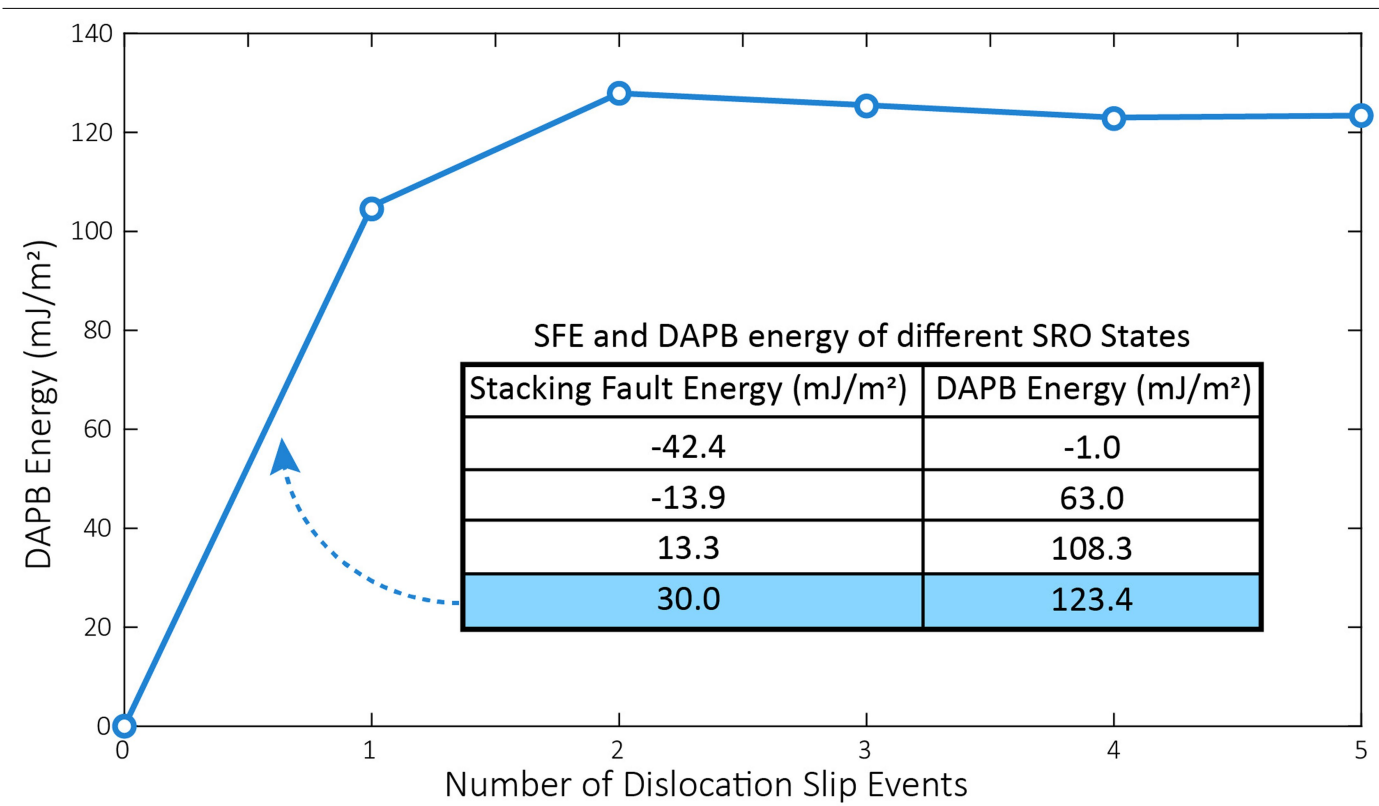


Extended Data Fig. 2 | Geometrical phase analysis strain mapping of a 1,000 °C aged sample and a water-quenched sample. a–d, 1,000 °C aged alloy; e–h, water-quenched sample. a, e, Drift-corrected HRSTEM images of the 1,000 °C aged sample and the water-quenched sample, respectively. The fast

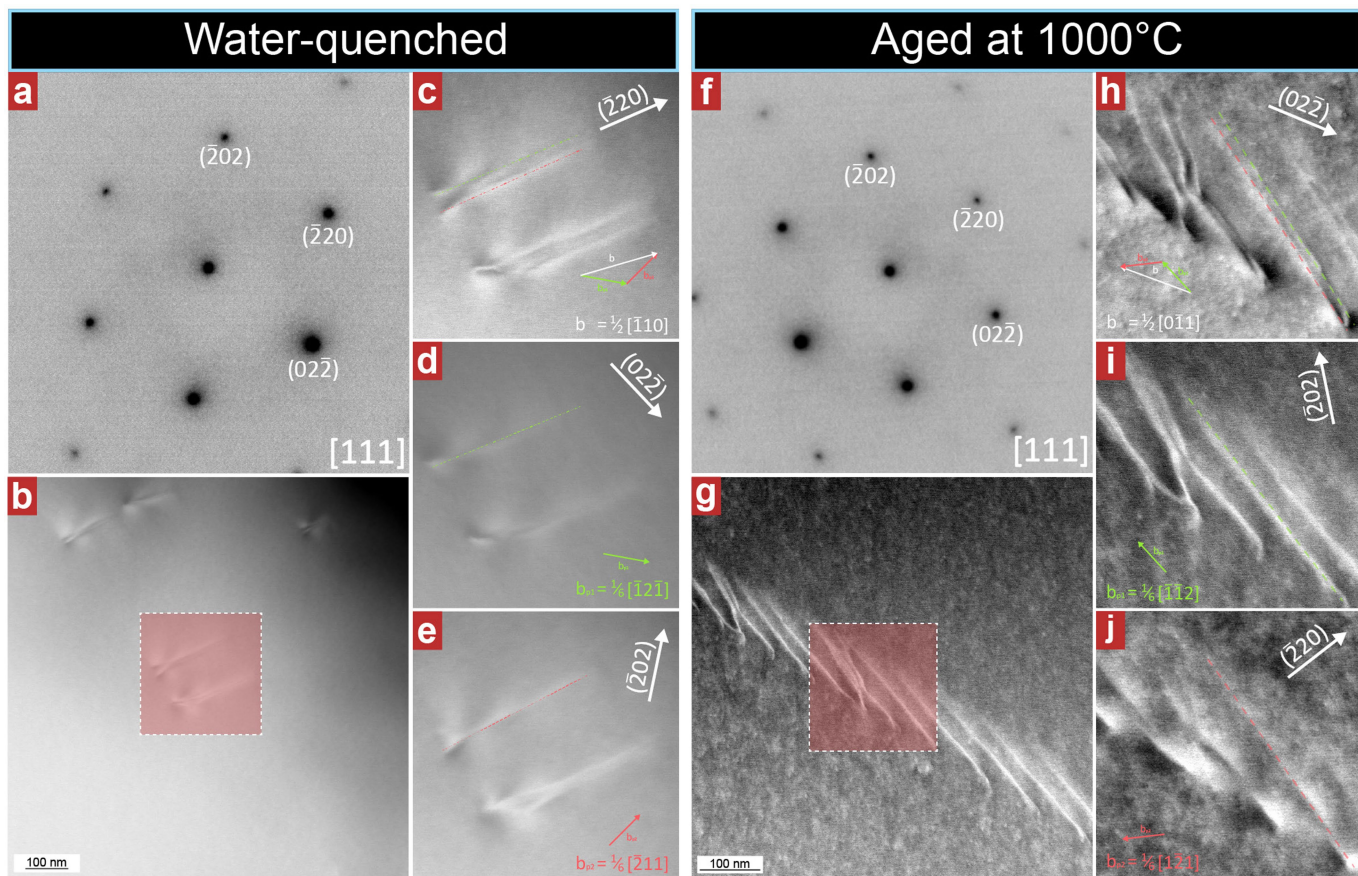
Fourier transformed images are shown inset. **b–d**, Strain maps of **a** showing nanometre-sized local fluctuations of strain (ϵ_{xx} , normal strain in the x direction; ϵ_{yy} , normal strain in the y direction). **f–h**, Strain maps of **e** showing similar but much weaker contrast of local strain.



Extended Data Fig. 3 | Results of X-ray diffraction experiments on a water-quenched sample and a 1,000 °C aged sample of the CrCoNi MEA. The indexed (111) and (200) peaks are marked. The lattice constants a are calculated on the basis of the 2θ angles of the identified peaks.



Extended Data Fig. 4 | Diffuse anti-phase boundary (DAPB) energy as a function of successive dislocation slip events from a calculated SRO model. The data in the inset table represent different states of SRO and the plot is from the state marked blue.



Extended Data Fig. 5 | Detailed 'g·b' analysis of partial dislocations in the CrCoNi MEA. a–e, Water-quenched sample; f–j, sample aged at 1,000 °C. a, f, Diffraction references showing the diffraction conditions (g** vectors) used for the analysis. b, g, Lower-magnification DC-STEM images of dislocations in**

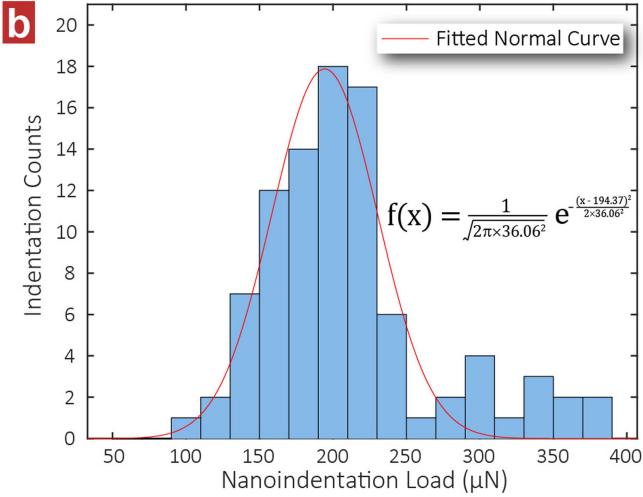
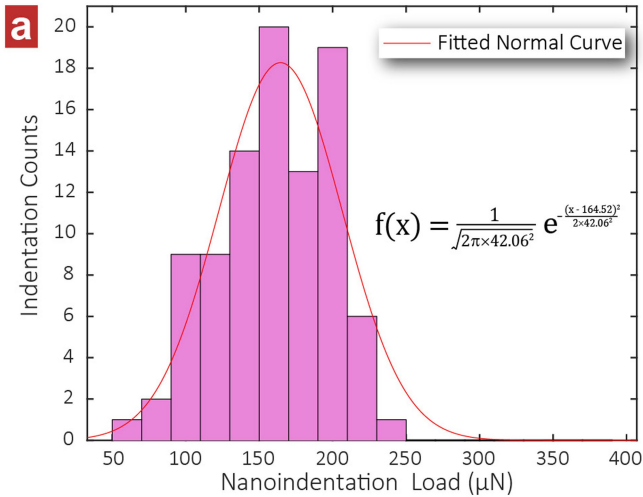
the water-quenched and aged samples, respectively. c–e, h–j, Two-beam DC-STEM images of the boxed areas in b and g, respectively; the Burgers vectors of the visible dislocations are noted on the images.

Water-quenched

Aged at 1000 °C

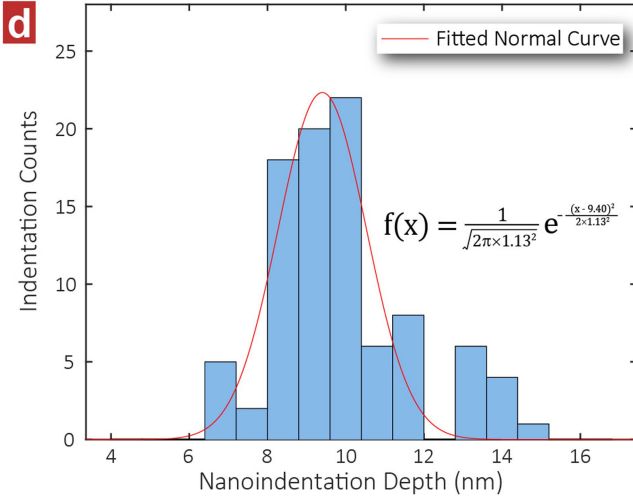
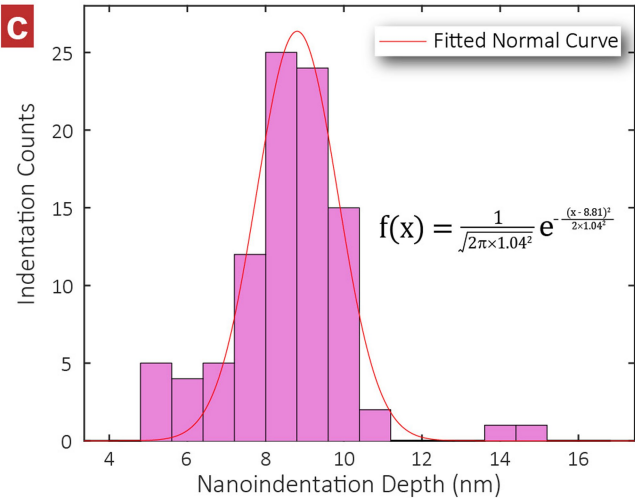
Distribution of Pop-in Load

Distribution of Pop-in Load



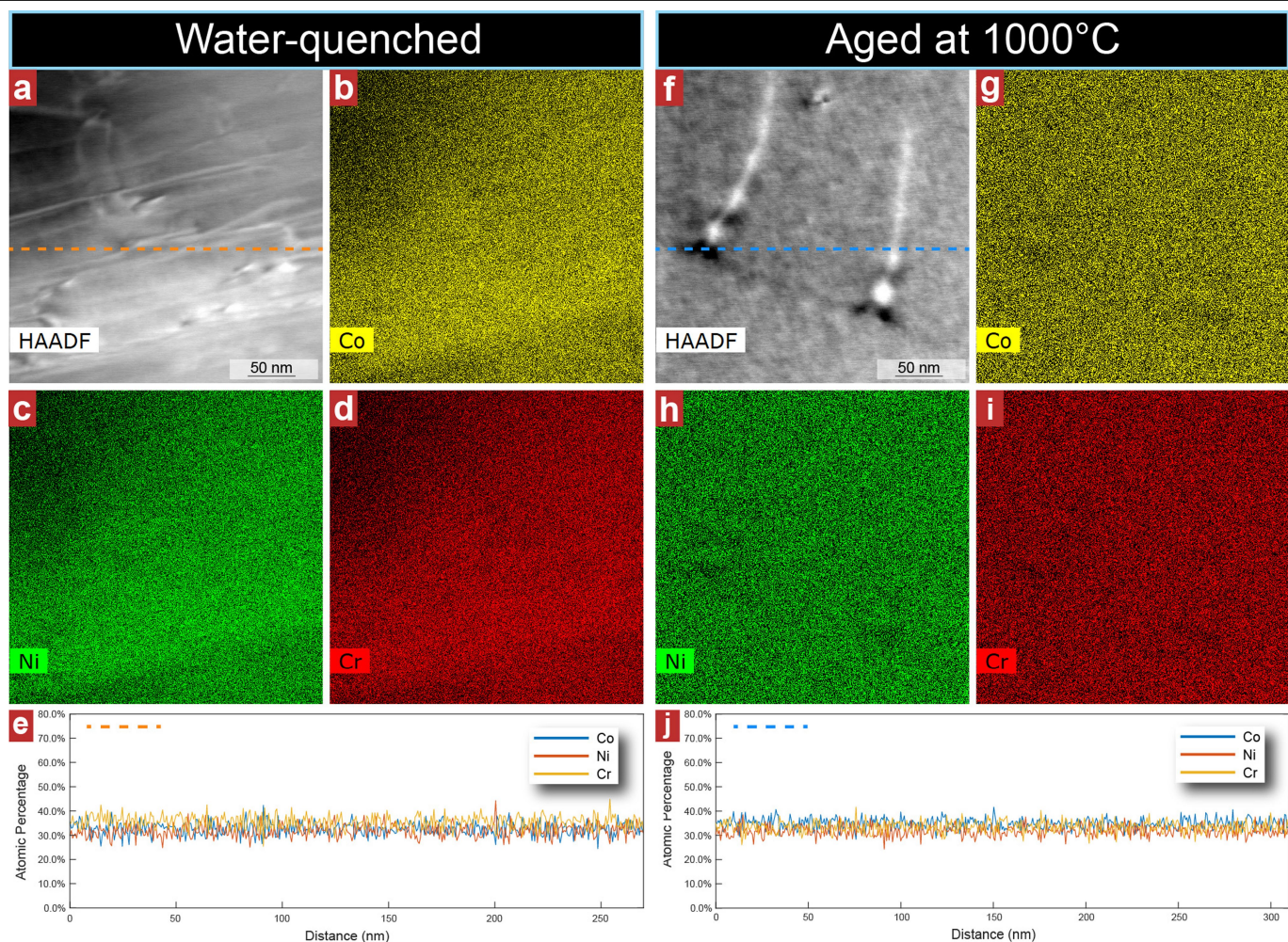
Distribution of Pop-in Depth

Distribution of Pop-in Depth



Extended Data Fig. 6 | Detailed statistical analysis of the pop-in events. Pop-in events were observed during nanoindentation tests (see Methods section ‘Nanoindentation experiments’ for details). **a, b**, Distribution of the pop-in load from water-quenched and 1,000 °C aged samples, respectively.

c, d, Distribution of the pop-in depth from water-quenched and 1,000 °C aged samples, respectively. The fitted normal distribution functions are listed in the panels. The results of numerical analysis are summarized in Extended Data Table 1.



Extended Data Fig. 7 | Element mapping of the water-quenched and aged CrCoNi samples using EDS. a–e, Water quenched sample; **f–j,** sample aged at 1,000 °C. **a, f,** Reference HAADF (high-angle annular dark field) images showing the regions of interest of a water-quenched sample and a 1,000 °C aged sample,

respectively. **b–d, g–i,** Element mapping of Co, Ni and Cr of the water-quenched sample and the 1,000 °C aged sample, respectively. **e, j,** Quantitative results of line scans of the water-quenched sample and the 1,000 °C aged sample, respectively. The line scan directions are marked by the dashed lines in **a** and **f**.

Extended Data Table 1 | Statistical results of SFE measurements and nanoindentation tests

		Water-Quenched	Aged at 1000 °C
Elastic Properties	Poisson's Ratio	0.29	0.28
	Young's Modulus (GPa)	229.9	231.0
	Shear Modulus (GPa)	89.1	90.2
Yield Strength	0.2% Offset Yield Strength (MPa)	205	255
Dislocation Dissociation	Partial Separation, (nm)	13.59 ± 2.64	6.44 ± 1.19
	SFE (mJ/m2)	8.18 ± 1.43	23.33 ± 4.31
Nanoindentation	Reduced Modulus (GPa)	181.76 ± 13.37	214.79 ± 18.49
	Indentation Hardness (GPa)	4.07 ± 0.23	4.37 ± 0.58
	Pop-in Load (µN)	164.52 ± 42.06	194.37 ± 36.06
	Pop-in Starting Displacement (nm)	8.81 ± 1.04	9.40 ± 1.13

Extended Data Table 2 | Detailed steps of the Gaussian template fitting process

Step	Description	Equation and comment
1	The standard deviation range of the Gaussian templates was set to 3 to 40 pixels (with a 0.1 interval) based on the pixel size of the DF image (0.056 nm/pixel).	
2	2-D Gaussian kernels with the same resolution as the DF images were constructed. The radius assumed for the SRO-enhanced domains was set to 1.3*sigma to best match the contrast observed in the DF image.	
3	To suppress the background noise during convolution, each Gaussian kernel was normalized by a larger Gaussian function to give a zero summation, using the expression at right, where G_{2D} is the normalized 2-D Gaussian template, σ is the varying standard deviation of the differently sized kernels, and x and y are the 2D coordinates from the kernel origin.	$G_{2D}(x, y, \sigma) = \frac{1}{2\pi\sigma^2} e^{-\frac{x^2+y^2}{2\sigma^2}} - \frac{1}{2\pi(1.5\sigma)^2} e^{-\frac{x^2+y^2}{2(1.5\sigma)^2}}$
4	The domain signals in the DF image were identified by a 2-D Gaussian Hough transform. For each kernel in the list, a 2-D convolution between the DF image and the kernel would be performed using the expression at right, where x is the DF image, h is the kernel and y is the result of the convolution.	$y[m, n] = \sum_j \sum_i x[i, j] \cdot h[m - i, n - j]$
5	After each convolution, pixels of the convolution result are compared to a data storing array; if the current pixel has a higher signal, the corresponding value would be updated in the data storing array. Another similarly sized array was used to store the associated kernel size of the highest signal.	
6	After the iterations, the pixel values were first filtered by the domain diameter range and the peak signal cutoff. Then the local peaks in the result array were identified if a pixel has a higher value than all of its eight neighbour pixels.	
7	The identified peaks were ordered and checked in a “brightest to dimmest” manner according to their pixel value. If dimmer peaks appear in the radius of a brighter peak, they would be deleted. This process is to eliminate overlapping entities.	
8	The remaining peaks were treated as identified domains.	
9	The diameter cutoff below 0.7 nm is set manually as this is already the size of the 1st nearest-neighbour shell of atoms in the MEA lattice, we do not regard anything below this value as an SRO domain.	

See Methods section ‘Energy-filtered TEM and SRO recognition’ for details.

Preparation of cyclohexene isotopologues and stereoisotopomers from benzene

<https://doi.org/10.1038/s41586-020-2268-y>

Received: 31 March 2019

Accepted: 11 March 2020

Published online: 20 May 2020

 Check for updates

Jacob A. Smith¹, Katy B. Wilson¹, Reilly E. Sonstrom¹, Patrick J. Kelleher¹, Kevin D. Welch¹, Emmitt K. Pert¹, Karl S. Westendorff¹, Diane A. Dickie¹, Xiaoping Wang², Brooks H. Pate¹ & W. Dean Harman^{1✉}

The hydrogen isotopes deuterium (D) and tritium (T) have become essential tools in chemistry, biology and medicine¹. Beyond their widespread use in spectroscopy, mass spectrometry and mechanistic and pharmacokinetic studies, there has been considerable interest in incorporating deuterium into drug molecules¹. Deutetrabenazine, a deuterated drug that is promising for the treatment of Huntington's disease², was recently approved by the United States' Food and Drug Administration. The deuterium kinetic isotope effect, which compares the rate of a chemical reaction for a compound with that for its deuterated counterpart, can be substantial^{1,3,4}. The strategic replacement of hydrogen with deuterium can affect both the rate of metabolism and the distribution of metabolites for a compound⁵, improving the efficacy and safety of a drug. The pharmacokinetics of a deuterated compound depends on the location(s) of deuterium. Although methods are available for deuterium incorporation at both early and late stages of the synthesis of a drug^{6,7}, these processes are often unselective and the stereoisotopic purity can be difficult to measure^{7,8}. Here we describe the preparation of stereoselectively deuterated building blocks for pharmaceutical research. As a proof of concept, we demonstrate a four-step conversion of benzene to cyclohexene with varying degrees of deuterium incorporation, via binding to a tungsten complex. Using different combinations of deuterated and proteated acid and hydride reagents, the deuterated positions on the cyclohexene ring can be controlled precisely. In total, 52 unique stereoisotopomers of cyclohexene are available, in the form of ten different isotopologues. This concept can be extended to prepare discrete stereoisotopomers of functionalized cyclohexenes. Such systematic methods for the preparation of pharmacologically active compounds as discrete stereoisotopomers could improve the pharmacological and toxicological properties of drugs and provide mechanistic information related to their distribution and metabolism in the body.

Typically, hydrogenation of benzene using D₂ gas leads to isotopologue mixtures of cyclohexane^{9–11}. However, Taube et al.¹² demonstrated that the complex [Os(NH₃)₅(η²-benzene)]²⁺ could be deuterated to form a single stereoisotopomer of [Os(NH₃)₅(η²-cyclohexene-d₄)]²⁺ using D₂ and a Pd/C catalyst. We posited that benzene bound in this manner could also be converted to cyclohexene using four well defined additions of two protons and two hydrides, passing through an η²-1,3-cyclohexadiene intermediate (Fig. 1). If these reactions could be performed regio- and stereoselectively, one could access a diverse set of isotopologues and even stereoisotopomers of cyclohexene using various combinations of proteated and deuterated reagents.

The dearomatization agent {Wtp(NO)(PMe₃)₃} is considerably more activating than its osmium predecessor¹³. Strong π-backbonding renders arene and diene complexes of this system highly nucleophilic

and resistant to substitution¹³. Furthermore, this system displays considerable electronic asymmetry, and the benzene complex Wtp(NO)(PMe₃)(η²-benzene) (**1**) can be prepared on a multi-gram scale¹⁴ and in enantioenriched form¹⁵. Treatment of an acetone-d₆ solution of **1** with diphenylammonium triflate (DPhAT; pK_a ≈ 0) at −30 °C affords its clean conversion to the η²-benzenium complex [Wtp(NO)(PMe₃)(η²-C₆H₇)] (OTf) (**2**; Fig. 2). Using chilled diethyl ether as a precipitating solvent, **2** can be isolated from dichloromethane in 86% yield (1.9 g). As an acetonitrile solution, the η²-benzenium complex **2** is moderately stable at room temperature but soon decomposes (half-life, *t*_{1/2} ≈ 6 min). At 0 °C, however, **2** exists in equilibrium with its diastereomer **3** in a 10:1 ratio (Fig. 2) and persists for three hours without substantial decomposition. The major isomer (**2**) is formed with the metal binding two internal carbons of the five-carbon π system, and with the newly formed

¹University of Virginia Department of Chemistry, Charlottesville, VA, USA. ²Neutron Scattering Division, Oak Ridge National Laboratory, Oak Ridge, TN, USA. ✉e-mail: wdh5z@virginia.edu

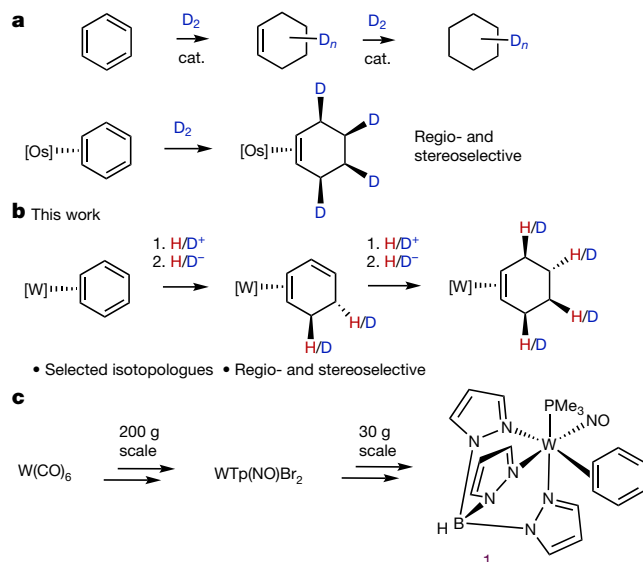


Fig. 1 | Methods for the deuteration of benzene. **a**, Existing methods for the selective deuteration of benzene can lead to over-reduction and a mixture of isotopologues. **b**, The current approach provides access to cyclohexene isotopologues and stereoisotopomers. **c**, The dearomatized benzene complex WTP(NO)(PMe₃)(η²-benzene).

*sp*³ carbon distal to the PMe₃ ligand. The minor isomer (**3**) is bound at a terminus of the π system with the *sp*³ carbon proximal to the phosphine. Proton nuclear magnetic resonance (NMR) data and density functional theory (DFT) calculations (Supplementary Information; Supplementary Figs. 1–3) of these η²-benzenium complexes (**2**, **3**) suggest that they are similar in structure to complexes of the form [WTP(NO)(PMe₃)(η²-allyl)]⁺ (ref. 16), where the allyl ligand is tightly bound to the metal through only two carbons. A third carbon, weakly associated to the metal, resembles a carbocation, and is indicated as such in the figures (Fig. 2). Combining cold solutions of **2** and tetrabutylammonium borohydride generates WTP(NO)(PMe₃)(η²-1,3-cyclohexadiene) exclusively (**4**). Despite the coexistence of the allyl conformer **3** in solution, the WTP(NO)(PMe₃)(η²-1,4-cyclohexadiene) complex (**8**) is undetected (Fig. 2) in the reaction mixture¹⁶. The η²-diene complex **4** is then treated with either DPhAT or HOTf/MeOH acids to generate the η²-allyl complex (**6**)¹⁶. When **6** is subjected to base, it deprotonates to form **5**, a stereoisomer of **4** (ref. 16), in which the uncoordinated double bond is now distal to the PMe₃ (ref. 16). Combining the allyl complex **6** with a hydride source produces the desired η²-cyclohexene complex **7** (67%). Crystals suitable for X-ray structure determinations are grown for complexes of cyclohexadiene **4**, allyl complex **6**, and cyclohexene **7**, and a rendering of these structures, along with key nuclear Overhauser effect (NOE) interactions are provided in Supplementary Information (Supplementary Fig. 4). Overlapping signals in the ¹H NMR spectrum of cyclohexene complex **7** precludes unambiguous stereochemical assignments of some of the ring proton signals.

By methylating the nitrosyl ligand of **7** (CH₃OTf) to generate [WTP(NOMe)(PMe₃)(η²-C₆H₁₀)]OTf (**9**)¹⁷, the chemical shifts of the cyclohexene ring separate to the point that each proton can be assigned with high confidence (Supplementary Information sections G and H). An X-ray structure determination of **9** provides conclusive evidence for methylation of the nitrosyl oxygen (Fig. 2), analogous to earlier literature reports¹⁸. Strong NOE interactions between the ring *endo* protons and the methylated nitrosyl ligand further facilitate these assignments, and quantitative NOE experiments support the stereochemical assignments of all diastereotopic protons on the cyclohexene ring (Supplementary Information section H).

Deuterium studies

With all hydrogen resonances for the methylated η²-cyclohexene complex **9** fully assigned, we investigated the regio- and stereochemical fidelity of the reaction sequence (Fig. 3). When the η²-benzenium complex **11** was prepared from **1** using [MeOD₂⁺]OTf, a loss of signal intensity was observed, corresponding to the methylene *endo* proton. This indicates that protonation of the η²-benzene occurs *syn* to the metal (Fig. 3). A complementary experiment was next performed, starting with the fully deuterated benzene complex, **17**, in which MeOH₂⁺ was used as the acid source. In this case, protonation led to a single broad proton resonance for the deuterated η²-benzenium complex **18**. This proton signal is -0.03 ppm upfield from its proteo counterpart, consistent with a primary H/D isotopic shift¹⁹. The *endo*-selective protonation of the benzene ligand in **1** is in stark contrast to the addition of carbon and heteroatom electrophiles, which have been observed to add *anti* to η²-arene and η²-diene ligands of tungsten complexes¹³. When η²-benzenium complexes **11** and **18** were treated with NaBD₄ or NaBH₄, respectively, the complementary cyclohexadiene complexes **12** and **19** were formed (Fig. 3). A comparison of NOESY data for all three isotopologues of the cyclohexadiene complex (**4**, **12**, **19**) confirmed that the proton delivered from the borohydride reagent was *anti* to the metal (Fig. 3). The cyclohexadiene complexes **12** and **19** were then taken forward to their π-allyl analogues **13**, **15** and **20** (Fig. 3). In contrast to protonation of the η²-benzene ligand of **1**, the acidic hydrogen was delivered predominantly *anti* to the metal (Fig. 3).

The resulting η²-allyl complexes (**13**, **15**, **20**) underwent a conformational change ('allyl shift') such that the second proton added became H^{6_{exo}} (conversion of **4** to **6**; Fig. 2), while the first proton added became H^{5_{endo}}. For allyl complexes **13** and **20**, full stereoselective protonation was achieved. However, with the preparation of **15** or **26** we experienced difficulties in achieving full deuterium incorporation, owing to an unusually large deuterium kinetic isotope effect (DKIE) (*k*_H/*k*_D ≈ 37 at -30 °C for the deuteration of **12** or **4**, where *k*_H and *k*_D are specific rate constants for protonation and deuteration, respectively). This DKIE was determined for **4** as the average value from three separate experiments in which **26** was generated from acidic solutions with different H/D ratios (Supplementary Information section K). This DKIE could be decreased by raising the temperature to 22 °C; however, this compromised the stereofidelity of the resulting deuterated product (**15**), with *endo* deuteration of the η²-diene **12**, which competed with *exo* deuteration. Consequently, stereoselective deuterium incorporation at the H^{6_{exo}} position of cyclohexene (that is, **16**, **33–35**, **41**, **44**, **49**, **51**; Fig. 3) could not be achieved above -75–80%. A similar outcome was observed when we tried to convert the *d*₆-isotopologue diene **19** to allyl **30**. Finally, treatment of **13**, **15** or **20** with a hydride or deuteride source again confirmed that the corresponding η²-cyclohexene products (**14**, **16**, **21**) are formed by nucleophilic addition *anti* to the metal (Fig. 3). Similarly to the 1,3-diene complex **4**, its isomer **5** undergoes *exo* protonation to form the allyl complex **24**. Remarkably, treatment of the 1,4-cyclohexadiene complex (**8**) with D⁺ (D₂NPh₂⁺ in MeOD) also undergoes direct *exo* protonation (Fig. 3), this time providing allyl **25**. The direct exogenous protonation of the unconjugated C=C bond in **8** appears to result in a carbocation that can be stabilized by the participation of the nitrosyl ligand, as revealed by DFT calculations. A subsequent [1,2]-hydride shift results in the formation of the allyl complex **25** (Supplementary Fig. 5). Unambiguous assignment of the deuterated hydrogen atom in **25** comes from its conversion to **9-d**₁ (via **39**; Fig. 3). To demonstrate regio- and stereocontrol of deuterium incorporation, additional deuterated isotopomers of the allyl complex were prepared from the monodeuterated dienes **22** and **23** and from the benzene-*d*₆-derived allyls **30** and **31** (Fig. 3). The allyl complexes **24–31** were then combined with deuteride or hydride to form 18 additional cyclohexene complexes, **32–46**, **49–51**. In principle, one can selectively make ten different isotopologues of the cyclohexene complex

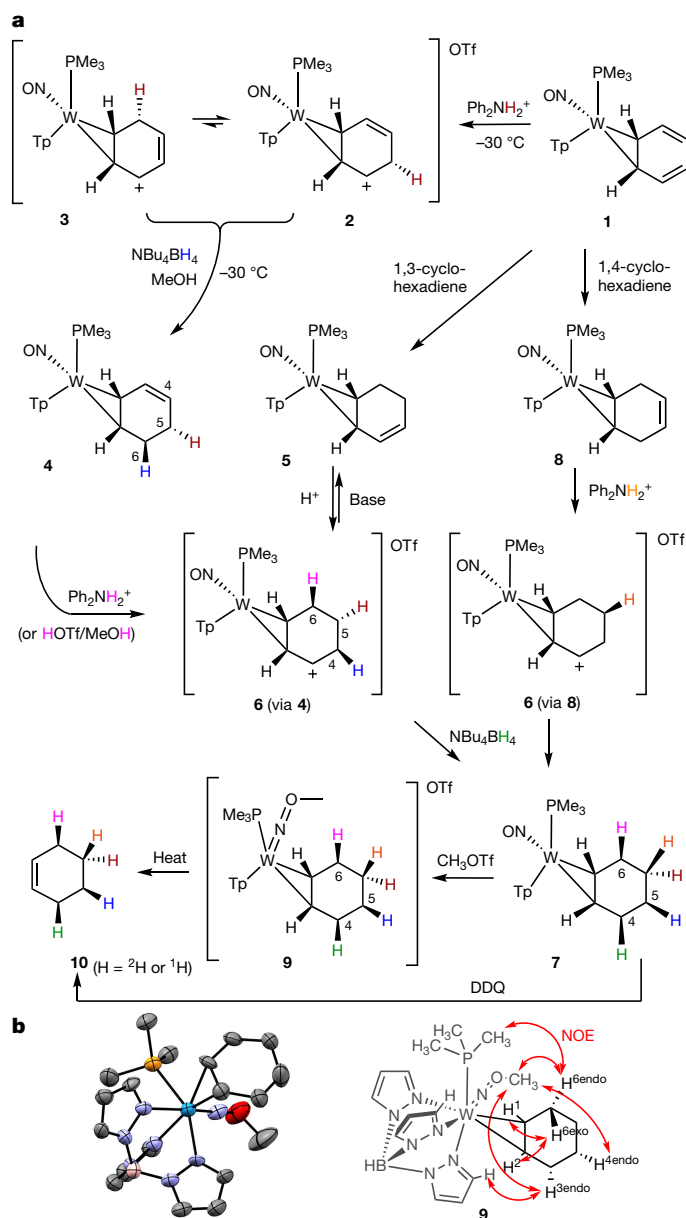


Fig. 2 | Formation of tungsten-bound cyclohexene from benzene.

a, Sequential reduction of benzene to cyclohexene bound to tungsten (by addition of ^1H or ^2H). **10** was confirmed by ^{13}C -NMR and MRR spectroscopy and **9** was confirmed by quantitative NOE. **b**, Solid-state molecular structure from a single-crystal X-ray diffraction study and relevant NOE interactions (red arrows) for the methylated cyclohexene complex **9** (Ph_2NH_2^+ as OTf salt).

using the procedures outlined above (d_0 – d_4 ; d_6 – d_{10}), eight of which (**7**, **16**, **32**–**38**) are reported herein.

Levels of isotopic purity for the cyclohexene ligand isotopologues were determined by recording high-resolution mass spectrometry (HRMS) data for the corresponding complexes as their methylated adducts (Fig. 2; **9**– d_n) to create a suitable cation for electrospray ionization mass analysis. Using the isotope envelope of **9**– d_0 as a reference (Supplementary Fig. 6), the isotopic purity of **7**, **16** and **32**–**38** (as converted to **9**– d_n) was estimated to be >90%, with the exception of **16** (79%), for which the high DKIE of the second protonation prevented complete deuteration at the $\text{H}^{6\text{exo}}$ position (see above). Finally, as a demonstration of how the $\{\text{Wtp}(\text{NO})(\text{PMe}_3)\}$ system precisely governs both the stereochemistry and the regiochemistry of protonation and hydride addition, a series of five monodeuterated (**32**, **39**–**42**), seven

dideuterated (**14**, **33**, **35**, **43**–**46**), and four trideuterated (**34**, **49**–**51**) isotopomers of the cyclohexene complex were prepared using these methods (Fig. 3).

Oxidation of the tungsten complex **7** with 2,3-dichloro-5,6-dicyano-*p*-benzoquinone (DDQ) releases the free cyclohexene (Fig. 2; **10**). Such action on **32**, **42**, **45** and **46** confirmed the expected regiochemistry of these d_1 and d_2 isotopomers of cyclohexene via ^{13}C NMR. Introduction of a single deuterium in 3-deuterocyclohexene or 4-deuterocyclohexene allows one to distinguish all six of the carbons in the ^{13}C NMR spectrum, owing to isotopic shifting of the now asymmetric cyclohexene carbons (Supplementary Fig. 7). Alternatively, solvent-free heating of various isotopologues of the methylated complex **9** induce the release of the cyclohexene ligand for analysis by molecular rotation resonance (MRR) spectroscopy (Supplementary Information section L)²⁰. These experiments determined that: (1) over-deuteration is exceedingly low (<2%). (2) The stereoselectivity is excellent when assessed by observation of undesired *cis/trans* isomers, which in the worst case is 22:1 and in other cases it is 40:1 or higher. (3) The dominant stereoisotopomers in all cases are those predicted by the ^1H NMR data. As a final check of the stereochemical assignments, the locations of the deuterium atoms were confirmed for complex **45** using neutron diffraction measurements (Supplementary Information section I; TOPAZ at Oak Ridge National Laboratory).

Mechanistic considerations

The reaction of **1** with D^+ to form **11** results in deuterium incorporation exclusively *endo* to the metal, but this does not conclusively show which carbon is initially protonated (Supplementary Fig. 8). Given that the *endo* proton of the benzene ligand in **1** completely preempts protonation from an exogenous acid (*exo*), we propose that the protonation must be concerted—that C–H bond formation is intramolecular and simultaneous with electronic changes at the metal—which could lower the activation barrier for this process relative to protonation by an external acid. Such a mechanism could occur via a hydride intermediate, but this seems sterically untenable. Instead, we propose a mechanism (Supplementary Information section M) in which the nitrosyl ligand is first protonated to form a hydroxylimido ligand analogous to that reported by Legzdins et al.²¹. This action is followed by a concerted proton transfer in which a gamma carbon of the benzene is protonated simultaneously with the release of electron density back into the tungsten through the NO group. The role of nitrosyl ligands in intramolecular proton transfer has been previously documented²². By contrast, the stereochemistry and kinetics of η^2 -diene protonation (for example, **4**; Fig. 2) indicates that the hydrogen is delivered exogenously, *anti* to the tungsten (Fig. 1). We speculate that whereas *endo* protonation may still be accessible for these 1,3-cyclohexadiene complexes, the less-delocalized diene ligand is probably more basic than its η^2 -benzene predecessor, and its direct *exo* protonation apparently preempts the purported *endo* mechanism at -30°C .

Transition-metal-promoted *endo* deuteration of benzene was observed in the η^4 -benzene complexes $\text{Cr}(\text{CO})_3(\eta^4\text{-benzene})^{2-}$ and $\text{Mn}(\text{CO})_3(\eta^4\text{-benzene})^-$ by Cooper et al.^{23,24}, and was proposed to occur via hydride intermediates^{23,24}. More recently, Chirik et al. explored the molybdenum-catalysed reduction of benzene and cyclohexadiene, with D_2 (g), which resulted in mixtures of isotopologues of cyclohexane¹¹. However, reduction of cyclohexene with D_2 produced a single *cis* isotopomer of 1,2-dideuterocyclohexane using the molybdenum catalyst.

The high stereoselectivity enabled by the tungsten system provides unprecedented control over the preparation of specific isotopologues and isotopomers of cyclohexene, starting from either benzene complex **1** or its deuterated analogue **17**, and using either proteated or deuterated sources of acids and hydrides (Supplementary Table 1). As an illustration, consider the d_2 isotopologue of the cyclohexene complex, **7**– d_2 . Given that the $\{\text{Wtp}(\text{NO})(\text{PMe}_3)\}$ system is available in

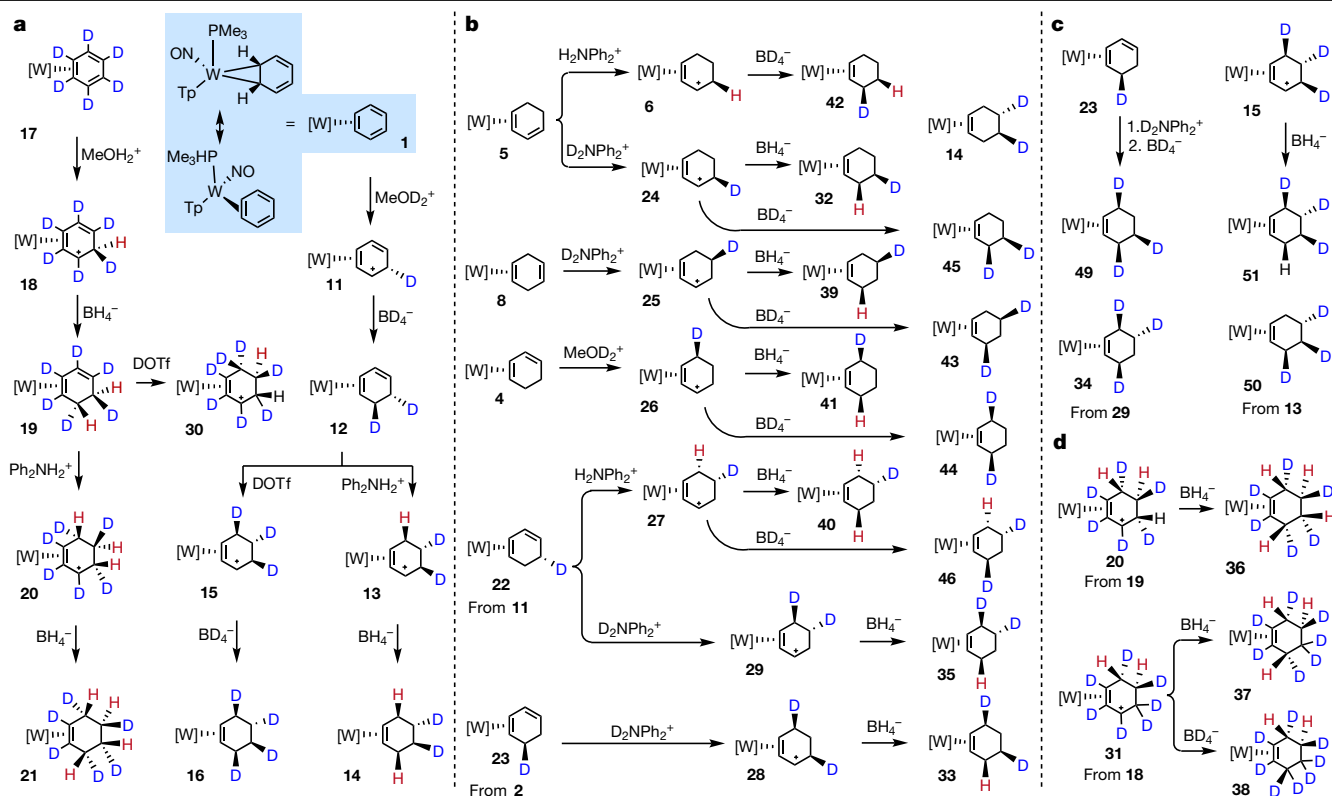


Fig. 3 | Synthesis of isotopologues and stereoisotopomers of the cyclohexene complex 7. a, Detailed syntheses of d_2 , d_4 and d_6 isotopologues. **b**, Synthesis of d_1 and d_2 isotopologues. **c**, Synthesis of d_3 isotopologues. **d**, Synthesis of d_6 , d_7 and d_8 isotopologues.

enantioenriched form¹⁵, one has access to 14 different isotopomers (individual enantiomers of **14**, **33**, **35**, **43–46**; Supplementary Table 2). The cyclohexene- d_2 ligand of these complexes, once removed from the metal by oxidative decomplexation, would be available as 11 individual isotopomers: both enantiomers of *cis*-3,4-, *trans*-3,4-, *cis*-3,5-, *trans*-3,5-, *trans*-4,5- and the meso compound *cis*-3,6-dideuterocyclohexene. Similarly, 11 distinct isotopomers of cyclohexene- d_3 should be available using this methodology starting from benzene- d_6 . Regarding cyclohexene complexes **7-d₃** and **7-d₇**, eight isotopomers of each would be available, and all 16 of these complexes would yield a unique, chiral cyclohexene (eight cyclohexene- d_3 and eight cyclohexene- d_7). In total (Supplementary Table 2), the methodology outlined herein could provide access to 52 unique isotopomers of cyclohexene derived from benzene and benzene- d_6 . For reference, the total number of isotopomers for cyclohexene is 528.

The ability of {Wtp(NO)(PMe₃)} to be optically resolved on a practical scale and to retain its stereochemical configuration, even when undergoing ligand displacement¹⁵, also makes it a valuable tool for determining the isotopic pattern of cyclohexene H/D isotopomers produced by other methods⁸. Consider, for example, a scenario in which an unknown isotopomer of cyclohexene- d_1 is combined with the resolved form of benzene complex (R)-**1** in solution and allowed to undergo ligand exchange. Even though the two faces of the cyclohexene ring will bind to tungsten with equal probability, the ¹H NMR spectrum will be unique for each of the five possible isotopomers (Supplementary Information section C; Supplementary Fig. 11). A similar approach could be taken for any cyclic alkene (for example, dehydropiperidines, pyrrolines, cyclopentenones) for which a ¹H NMR spectrum of a fully protected species can be fully assigned (see above).

Deuterated building blocks for the MedChem database

The development of deutetrabenazine, is considered by many as a prelude to a new generation of medicines and therapies that incorporate

deuterium into the active pharmaceutical ingredient². Given that each stereoisotopomer of a biologically active substance will have its own unique pharmacokinetic profile, the ability to stereoselectively deuterate cyclohexene or other MedChem building blocks could enable the development of new probes, fragment libraries and leads for medicinal chemists, as well as provide a new tool for organic and organometallic mechanistic studies. Cyclohexene can be readily converted into perhydroindoles²⁵, perhydroisoquinolines²⁶ and azepines²⁷. However, the inability to chemically differentiate the two alkene carbons or the enantioface of the deuterated cyclohexene limits its potential. Nevertheless, by replacing the benzene ligand in Fig. 2 with a substituted benzene, or by using a non-hydrogenic nucleophile in the conversion of **6** to **7** (Fig. 2), one can envision a series of 3-substituted cyclohexenes with highly defined isotopic patterns. As proof of concept, we prepared the α,α,α -trifluorotoluene complex Wtp(NO)(PMe₃)(η^2 -CF₃Ph) (ref. ²⁸), which can be elaborated into a 3-(trifluoromethyl)cyclohexene complex (**47**) analogous to the cyclohexene complex **7** (Supplementary Fig. 14). Liberation of the cyclohexene from {Wtp(NO)(PMe₃)} can be accomplished by a one-electron oxidant such as DDQ, Fe(III) or NOPF₆ in yields of 70–75% (ref. ²⁸). Oxidation of **47** would generate a cyclohexene that has been previously shown to undergo diastereoselective epoxidation, and would therefore be an attractive building block for medicinal chemistry²⁹. Repeating the synthesis of **47** with deuteride in the final step yields the *cis*-6-deutero-3-(trifluoromethyl)cyclohexene complex **52** in 95% yield. Various other isotopologues of **47** and **52** were also prepared (**47**, **52**, **53**, **54**), and the reaction pattern was found to be similar to that observed for benzene. The prepared compounds are summarized in Fig. 4, with synthetic details provided in Supplementary Information section B. Notably, in the syntheses of **47**, **52**, **53** and **54**, protonation at the carbon bearing the CF₃ group ultimately occurs *endo* to the metal, allowing the CF₃ group to assume an *exo* stereochemistry. However, if the purported diene intermediate is protonated under kinetic control, *exo* protonation forces the CF₃ group *endo*, and the result after a second hydride reduction is the cyclohexene complex

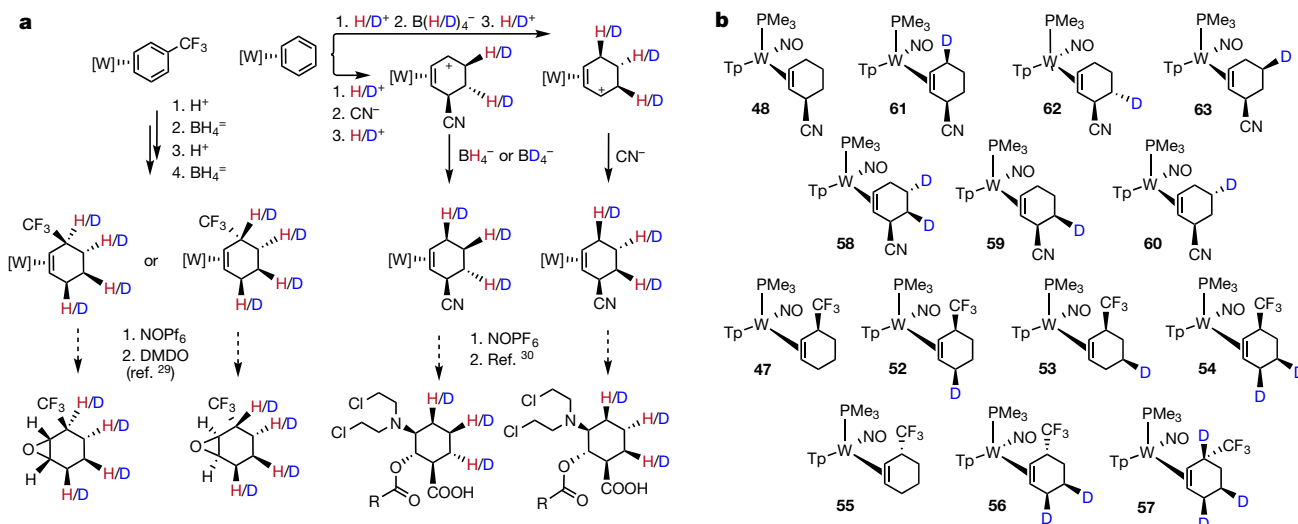


Fig. 4 | Examples of functionalized cyclohexene isotopomer complexes. **a**, Synthesis (speculated) of functionalized cyclohexenes; ref. ²⁹ describes a single-step reaction with DMDO (70%); ref. ³⁰ describes a nine-step synthesis that includes: (i) HCl/H₂O, (ii) isobutylene, (iii) mCPBA, (iv) NaCN, (v) acetyl

chloride, (vi) H₂/C, (vii) ethylene oxide, (viii) TsCl and (ix) TFA. **b**, Examples of chemo- and stereoselectively deuterated cyclohexene complexes with CF_3 and CN groups on the cyclohexene. See Supplementary Information section B for the full synthetic details of **47**, **48**, **52–63**.

55. By exploiting this reactivity feature, we were able to prepare other isotopologues of **55** with inversion of the stereocentre bearing the $-CF_3$ substituent (Fig. 4, **56**, **57**; Supplementary Fig. 14).

As further demonstration of the ability of this methodology to selectively prepare isotopomers of functionalized cyclohexenes, we prepared the tungsten complex of *cis,trans*-3-cyano-4,5-dideuterocyclohexene (**58**) by the addition of cyanide to the allyl intermediate **13** (57%; diastereometric ratio >98%; Fig. 4.). Other d_1 -isotopologues were also prepared (Supplementary Fig. 15), and their stereochemistry could again be controlled with the sequence of nucleophiles. For example, **58**, **59** and **60** could be prepared by generating the appropriate isotopologue of the tungsten-allyl complex and then treating with NaCN (Supplementary Fig. 16). Conversely, treating the benzenium **2** with NaCN leads to a cyano-substituted cyclohexadiene that can be subsequently combined with acid and a hydride source to generate other cyclohexene isotopomers (**61–63**; Supplementary Information section B). 3-cyanocyclohexene (proteo form) has been previously used as a precursor to cytotoxic mustards that are of interest in cancer research³⁰. Allyl-substituted cyclohexenes theoretically exist as 1,024 different H/D isotopomers (512 for each enantiomer). Using the tungsten dearomatization methodology, the CF_3 - and CN -substituted cyclohexenes are accessible as **64** and **60** unique isotopomers, respectively. We further note that a full range of both carbon and nitrogen nucleophiles has now been demonstrated to add to tungsten benzenium and allyl tungsten complexes³¹, which demonstrates the broad scope of compounds that can now be prepared as various deuterioisotopomers.

Online content

Any methods, additional references, Nature Research reporting summaries, source data, extended data, supplementary information, acknowledgements, peer review information; details of author contributions and competing interests; and statements of data and code availability are available at <https://doi.org/10.1038/s41586-020-2268-y>.

- Gant, T. G. Using deuterium in drug discovery: leaving the label in the drug. *J. Med. Chem.* **57**, 3595–3611 (2014).
- Dean, M. & Sung, V. W. Review of deutetrabenazine: a novel treatment for chorea associated with Huntington's disease. *Drug Des. Devel. Ther.* **12**, 313–319 (2018).

- Thibblin, A. & Ahlberg, P. Reaction branching and extreme kinetic isotope effects in the study of reaction mechanisms. *Chem. Soc. Rev.* **18**, 209–224 (1989).
- Thibblin, A. Unusually large kinetic deuterium isotope effects on oxidation reactions. 1. the mechanism of hydroxide-catalysed permanganate oxidation of PhCD(CF₃)OH and PhCD(CH₃)OH in water. *J. Phys. Org. Chem.* **8**, 186–190 (1995).
- Nelson, S. D. & Trager, W. F. The use of deuterium isotope effects to probe the active site properties, mechanism, of cytochrome P450-catalyzed reactions, and mechanisms of metabolically dependent toxicity. *Drug Metab. Dispos.* **31**, 1481–1497 (2003).
- Loh, Y. Y. et al. Photoredox-catalyzed deuteration and tritiation of pharmaceutical compounds. *Science* **358**, 1182–1187 (2017).
- Pony Yu, R., Hesk, D., Rivera, N., Pelczar, I. & Chirik, P. J. Iron-catalysed tritiation of pharmaceuticals. *Nature* **529**, 195–199 (2016).
- Baldwin, J. E., Kiemle, D. J. & Kostikov, A. P. Quantitative analyses of stereoisomeric 3,4- d_2 -cyclohexenes in the presence of 3,6- d_2 -cyclohexenes. *J. Org. Chem.* **74**, 3866–3874 (2009).
- Eisen, M. S. & Marks, T. J. Supported organoactinide complexes as heterogeneous catalysts. A kinetic and mechanistic study of facile arene hydrogenation. *J. Am. Chem. Soc.* **114**, 10358–10368 (1992).
- Jones, R. A. & Seeberger, M. H. Synthesis of polymer-supported transition metal catalysts via phosphido linkages: heterogeneous catalysts for the hydrogenation of aromatic compounds under mild conditions. *J. Chem. Soc. Chem. Commun.* 373–374 (1985).
- Joannou, M. V., Bezdek, M. J. & Chirik, P. J. Pyridine(diimine) molybdenum-catalyzed hydrogenation of arenes and hindered olefins: insights into precatalyst activation and deactivation pathways. *ACS Catal.* **8**, 5276–5285 (2018).
- Harman, W. D. & Taube, H. The selective hydrogenation of benzene to cyclohexene on pentaammineosmium(II). *J. Am. Chem. Soc.* **110**, 7906–7907 (1988).
- Liebov, B. K. & Harman, W. D. Group 6 dihapto-coordinate dearomatization agents for organic synthesis. *Chem. Rev.* **117**, 13721–13755 (2017).
- Welch, K. D. et al. Large-scale syntheses of several synthons to the dearomatization agent {TpW(NO)(PMe₃)₃} and convenient spectroscopic tools for product analysis. *Organometallics* **26**, 2791–2794 (2007).
- Lankenau, A. W. et al. Enantioenrichment of a tungsten dearomatization agent utilizing chiral acids. *J. Am. Chem. Soc.* **137**, 3649–3655 (2015).
- Harrison, D. P. et al. Hyperdistorted tungsten allyl complexes and their stereoselective deprotonation to form dihapto-coordinated dienes. *Organometallics* **30**, 2587–2597 (2011).
- Lis, E. C. et al. The uncommon reactivity of dihapto-coordinated nitrile, ketone, and alkene ligands when bound to a powerful π -base. *Organometallics* **25**, 5051–5058 (2006).
- Arashiba, K. et al. Electrophilic O-methylation of a terminal nitrosyl ligand attained by an early–late heterobimetallic effect. *Organometallics* **25**, 560–562 (2006).
- Jamison, C. J. Isotope effects on chemical shifts and coupling constants. *eMagRes* <https://doi.org/10.1002/9780470034590.emrstm0251> (2007).
- Pérez, C. et al. Broadband Fourier transform rotational spectroscopy for structure determination: the water heptamer. *Chem. Phys. Lett.* **571**, 1–15 (2013).
- Sharp, W. B., Legzdins, P. & Patrick, B. O. O-Protonation of a terminal nitrosyl group to form an η^1 -hydroxylimido ligand. *J. Am. Chem. Soc.* **123**, 8143–8144 (2001).
- Llamazares, A., Schmalke, H. W. & Berke, H. Ligand-assisted heterolytic activation of hydrogen and silanes mediated by nitrosyl rhenium complexes. *Organometallics* **20**, 5277–5288 (2001).

23. Leong, V. S. & Cooper, N. J. Electrophilic activation of benzene in $[\text{Cr}(\eta^4\text{-C}_6\text{H}_6)(\text{CO})_3]^{2-}$. *J. Am. Chem. Soc.* **110**, 2644–2646 (1988).
24. Thompson, R. L., Lee, S., Rheingold, A. L. & Cooper, N. J. Reductive activation of the coordinated benzene in manganese complex $[\text{Mn}(\eta^5\text{-C}_6\text{H}_6)(\text{CO})_3]^+$: synthesis and characterization of the η^4 -naphthalene complex $\text{PPN}[\text{Mn}(\eta^4\text{-C}_{10}\text{H}_8)(\text{CO})_3]$. *Organometallics* **10**, 1657–1659 (1991).
25. Ungureanu, I., Klotz, P., Mann, A. Phenylaziridine as a masked 1,3 dipole in reactions with nonactivated alkenes. *Angew. Chem.* **39**, 4615–4617 (1991).
26. Sarkar, N., Banerjee, A. & Nelson, S. G. $[4 + 2]$ Cycloadditions of *N*-alkenyl iminium ions: structurally complex heterocycles from a three-component Diels–Alder reaction sequence. *J. Am. Chem. Soc.* **130**, 9222–9223 (2008).
27. Shim, S. C., Doh, C. H., Kim, T. J., Lee, H. K. & Kim, K. D. A new and convenient synthesis of *N*-substituted perhydroazepines from adipaldehyde and primary amines with tetracarbonylhydrideferrate, $\text{HFe}(\text{CO})_4^-$, as a selective reducing agent. *J. Heterocycl. Chem.* **25**, 1383–1385 (1988).
28. Wilson, K. B. et al. Sequential tandem addition to a tungsten–trifluorotoluene complex: a versatile method for the preparation of highly functionalized trifluoromethylated cyclohexenes. *J. Am. Chem. Soc.* **139**, 11401–11412 (2017).
29. Murray, R. W., Singh, M., Williams, B. L. & Moncrieff, H. M. Diastereoselectivity in the epoxidation of substituted cyclohexenes by dimethyldioxirane. *J. Org. Chem.* **61**, 1830–1841 (1996).
30. Leiris, S., Lucas, M., Dupuy d’Angeac, A. & Morère, A. Synthesis and biological evaluation of cyclic nitrogen mustards based on carnitine framework. *Eur. J. Med. Chem.* **45**, 4140–4148 (2010).
31. Wilson, K. B. et al. Highly functionalized cyclohexenes derived from benzene: sequential tandem addition reactions promoted by tungsten. *J. Org. Chem.* **84**, 6094–6116 (2019).

Publisher’s note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

© The Author(s), under exclusive licence to Springer Nature Limited 2020

Data availability

All data are available in the main text and Supplementary Information, including NMR spectra, experimental details, crystallographic information, DFT calculations, rotational spectroscopy and HRMS data. Supplementary crystallographic data for this paper (**4**, **7**, **9** (X-ray) and **45** (neutron)) can be obtained from the Cambridge Crystallographic Data Centre at www.ccdc.cam.ac.uk/structures (CCDC 1885723-1885725 and 1972890).

Acknowledgements We acknowledge the assistance of E. Ashcraft in collecting HRMS data. The work was funded by the National Institutes of Health (1R01GM132205-01) and the University of Virginia. The single-crystal neutron diffraction experiment performed on TOPAZ using resources at the Spallation Neutron Source, a DOE Office of Science User Facility operated by the Oak Ridge National Laboratory, under contract number DE-AC05-00OR22725 with UT-Battelle, LLC.

Author contributions W.D.H., J.A.S. and K.B.W. conceived the project. K.D.W., W.D.H. and J.A.S. designed the experiments. J.A.S. and K.S.W. prepared the samples and collected NMR and HRMS data. D.A.D. carried out X-ray molecular structure determinations. B.H.P., P.J.K. and R.E.S. conceived and ran the rotational spectroscopy experiments. E.K.P. and K.S.W. carried out DFT calculations. J.A.S. and W.D.H. wrote the manuscript. X.W. collected and processed neutron diffraction data.

Competing interests The authors declare no competing interests.

Additional information

Supplementary information is available for this paper at <https://doi.org/10.1038/s41586-020-2268-y>.

Correspondence and requests for materials should be addressed to W.D.H.

Peer review information *Nature* thanks David Hesk and the other, anonymous, reviewer(s) for their contribution to the peer review of this work.

Reprints and permissions information is available at <http://www.nature.com/reprints>.

Patterns and trends of Northern Hemisphere snow mass from 1980 to 2018

<https://doi.org/10.1038/s41586-020-2258-0>

Received: 6 June 2019

Accepted: 25 February 2020

Published online: 20 May 2020

 Check for updates

Jouni Pulliainen¹, Kari Luojus¹, Chris Derksen², Lawrence Mudryk², Juha Lemmetyinen¹, Miia Salminen¹, Jaakko Ilkonen¹, Matias Takala¹, Juval Cohen¹, Tuomo Smolander¹ & Johannes Norberg^{1✉}

Warming surface temperatures have driven a substantial reduction in the extent and duration of Northern Hemisphere snow cover^{1–3}. These changes in snow cover affect Earth's climate system via the surface energy budget, and influence freshwater resources across a large proportion of the Northern Hemisphere^{4–6}. In contrast to snow extent, reliable quantitative knowledge on seasonal snow mass and its trend is lacking^{7–9}. Here we use the new GlobSnow 3.0 dataset to show that the 1980–2018 annual maximum snow mass in the Northern Hemisphere was, on average, $3,062 \pm 35$ billion tonnes (gigatonnes). Our quantification is for March (the month that most closely corresponds to peak snow mass), covers non-alpine regions above 40° N and, crucially, includes a bias correction based on in-field snow observations. We compare our GlobSnow 3.0 estimates with three independent estimates of snow mass, each with and without the bias correction. Across the four datasets, the bias correction decreased the range from 2,433–3,380 gigatonnes (mean 2,867) to 2,846–3,062 gigatonnes (mean 2,938)—a reduction in uncertainty from 33% to 7.4%. On the basis of our bias-corrected GlobSnow 3.0 estimates, we find different continental trends over the 39-year satellite record. For example, snow mass decreased by 46 gigatonnes per decade across North America but had a negligible trend across Eurasia; both continents exhibit high regional variability. Our results enable a better estimation of the role of seasonal snow mass in Earth's energy, water and carbon budgets.

How much water is stored in seasonal snow? The simplicity of the question belies the difficulties we have in answering it. Snow cover is sensitive to changes in precipitation and temperature—the drivers of snowfall events—expressed at scales of both short-term weather phenomena and long-term monthly-to-seasonal climate^{1,5,6,10}. This leads to a high spatial and temporal variability in the distribution of snow mass. Meltwater from seasonal snow recharges essential freshwater resources, stored as surface water in rivers and lakes, soil moisture and groundwater. Across most of the Northern Hemisphere land regions, the annual runoff is dominated by snowmelt⁴. With a warming climate, there are regionally and temporally nonuniform shifts in snowmelt, runoff and precipitation patterns^{10,11}. Besides providing water, seasonal snow cover has a key role in the global energy cycle by cooling near-surface temperatures¹². Anomalies in snow mass influence snow-dependent ecosystems, populations and economic activities even in downstream snow-free regions^{6,10,13}. Furthermore, by insulating the underlying soil and influencing vegetation phenology and fire risk, seasonal snow has a substantial impact on carbon-cycle processes, in particular by affecting the magnitude of carbon fluxes at the regions of seasonal frozen soil and permafrost^{14–16}.

Limitations of snow mass assessments

Although terrestrial snow is a World Meteorological Organization (WMO) Global Climate Observing System (GCOS) Essential Climate

Variable (ECV), we still lack simple baseline knowledge of the Northern Hemisphere mass of seasonal snow^{8,13,17}. In areas in which the snow mass over watersheds at the end of winter is poorly known, predictions of water-cycle processes are highly uncertain^{4,6}. Thus, accurate knowledge of snow mass is required in order to close the water cycle, provide fundamental estimates of freshwater availability, and enable accurate initialization of hydrological models, which are important tools to minimize the impact of potentially hazardous flood events^{4,6}. Further, in operational weather prediction, inadequate information on snow mass decreases the accuracy of weather forecasts at local, regional and global scales owing to land–atmosphere feedbacks¹⁸. Climate modelling and analysis demand improved quantification of the snow–atmosphere interaction through more accurate observations and modelling of energy and mass exchanges, including the contribution of snow to freshwater runoff.

Previous assessments of seasonal hemispheric snow mass—based on satellite data, ground-based networks and land surface models driven by historical atmospheric reanalyses—provide highly variable estimates, owing to both spatiotemporal inaccuracies in driving data and simplified descriptions of physical snow processes⁸. Analysis of various data sources spanning the 1981–2010 period show a relative uncertainty of approximately 50% in climatological hemispheric peak snow mass⁸, with even higher uncertainties for mountain regions¹⁷. Earlier investigations suggest a decline in continental-scale snow mass during the past decades, even though spatial trend patterns are highly variable between different

¹Finnish Meteorological Institute, Helsinki, Finland. ²Climate Research Division, Environment and Climate Change Canada, Toronto, Ontario, Canada. ✉e-mail: jouni.pulliainen@fmi.fi

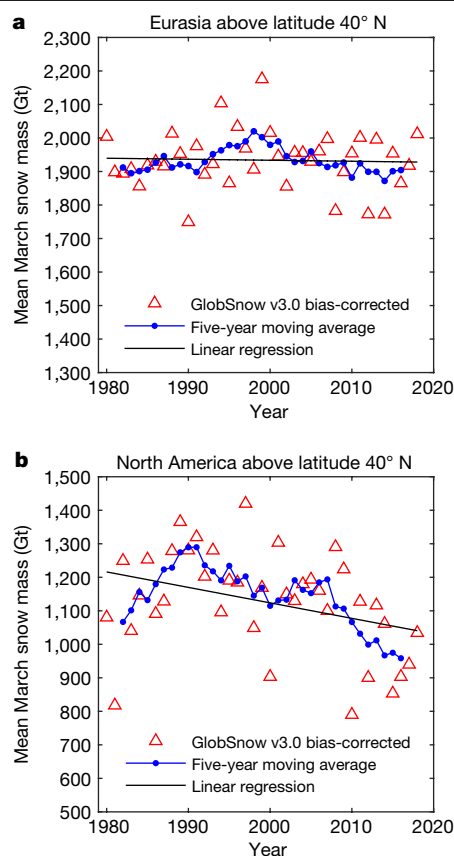


Fig. 1 | March annual snow mass and its trend. GlobSnow v3.0 bias-corrected estimates for the period of 39 years (1980–2018). **a**, Eurasia and **b**, North America.

products^{7–9,19}. A thorough validation of obtained trends and snow mass estimates has not been possible previously owing to a lack of the extensive independent hemispheric in situ reference datasets required for evaluation.

Satellite instruments operating at visible and microwave frequencies provide an effective tool with which to obtain information on the regional and global patterns of snow cover extent^{3,20}. However, the estimation of snow mass—often given by gridded snow water equivalent values (SWE, the depth of water released by instantaneous snow melt)—has proven to be problematic. Attempts to assess the mass of global seasonal snow cover through the use of spaceborne microwave radiometry have been carried out since the launch of the Scanning Multi-channel Microwave Radiometer (SMMR) instrument in 1978 (ref. ²¹). SMMR and the succeeding passive microwave instruments provide a daily data record of microwave brightness temperatures (T_b) observed at a coarse spatial resolution (tens of kilometres). T_b measurements alone can yield reasonable SWE estimates for a certain region and/or winter, but lack sufficient consistency in performance for global multiyear applications^{9,20,22–24}. Spaceborne T_b observations are influenced by various characteristics of the land surface (soil and vegetation), the atmosphere and the snowpack itself^{25,26}. For example, the layered structure of a natural snowpack and the snow microstructure (snow grain size) have strong effects on the propagation of microwave radiation within snow, which causes severe problems for SWE retrieval algorithms because these characteristics vary both spatially and temporally, without a simple correlation to the variations in SWE^{25,27}.

To overcome the shortcomings of stand-alone algorithms using microwave radiometry as a sole information source, an approach was developed that combines T_b data with ground-based measurements of snow depth from the global network of synoptic weather stations²⁴. This approach was selected as the baseline method for the Global Snow Monitoring for

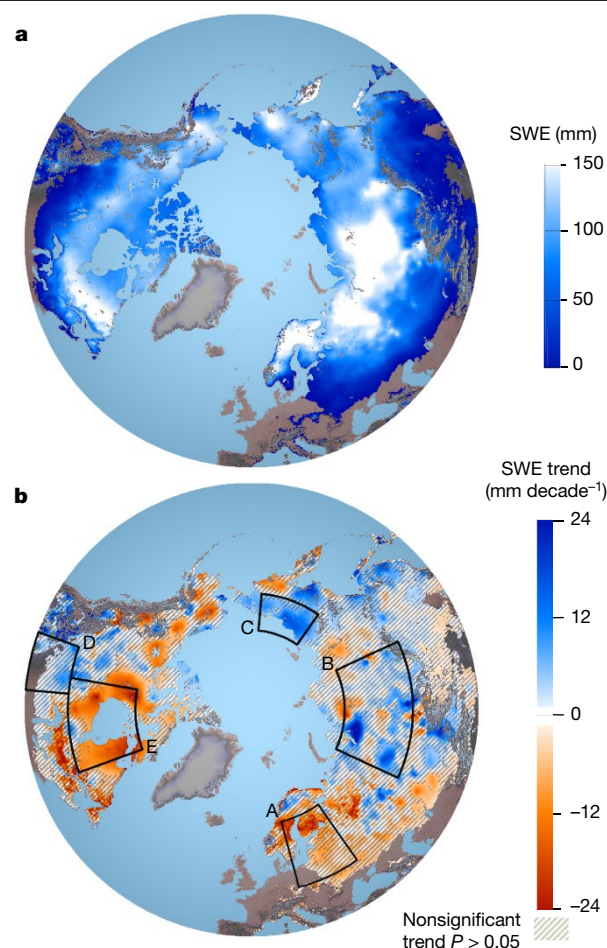


Fig. 2 | Distribution and decadal trend of mean March hemispheric snow mass. **a**, Average snow distribution during 1980–2018, given in terms of SWE (mm). **b**, Decadal SWE trend for the period 1980–2018. Hatches indicate areas with a statistically nonsignificant trend ($P > 0.05$). The illustrated snow region (blue to white areas in **a**) includes areas in which the estimated March SWE exceeded 5 mm (snow depth greater than around 2–4 cm) on average during the period 1980–2018. Mountain regions with high topographic variability are excluded. A–E indicate our dedicated areas for investigating regional trends.

Climate Research (GlobSnow) initiative of the European Space Agency (ESA), which produced a climate data record (CDR) of global snow mass evolution²⁸. Considering that, at present, SWE estimates that rely solely on passive microwave observations exhibit large errors when compared with independent reference data^{22–24}, and that global reanalysis-driven products do not as yet assimilate surface snow observations, the GlobSnow method provides the most reliable existing observation-based (satellite plus in situ) estimates of the hemispheric-scale snow mass (SWE). Improved accuracy can be obtained by fusing together the satellite data and distributed in situ observations of snow characteristics using a maximum likelihood optimization, although the method has limitations, particularly concerning the estimation of high values of SWE²⁸ (Extended Data Fig. 1). While some studies indicate that land surface model estimates can approach the GlobSnow method regarding accuracy²⁹, a basic goal of the GlobSnow CDR was the provision of SWE from observations (satellite and in situ) independently of climate or reanalysis model estimates.

Reduction of snow mass uncertainty

Here, we enhance the GlobSnow approach by developing a bias-correction method, which combines the GlobSnow v3.0 SWE

Table 1 | March snow mass and its global, continental and regional decadal trends

Region above 40° N: non-alpine snow area (minimum/average/maximum) (×10 ⁶ km ²)	Mean monthly snow mass 1980–2018 (with error bounds) ^a (Gt)	Average SWE ^b (mm)	Snow mass trend and 95% confidence interval (Gt decade ⁻¹)	P-value of trend line
Northern Hemisphere (29/32/35)				
GlobSnow v3.0	3,062 ± 5	97	−49 ± 49	0.048
Average (range), all four products ^c	2,938 (2,846–3,062)	(90–97)		
Average (range), all without bias correction	2,867 (2,433–3,380)	(77–107)		
Alpine areas, average (range) ^d	709 (482–937)	(96–187)		
Eurasia (19/21/23)				
GlobSnow v3.0	1,934 ± 4	91	−3 ± 25 ^e	0.809 ^e
Average (range), all four products ^c	1,836 (1,736–1,934)	(81–91)		
Average (range), all without bias correction	1,895 (1,610–2,229)	(76–106)		
Alpine areas, average (range) ^d	225 (178–298)	(59–99)		
North America (9/11/12)				
GlobSnow v3.0	1,128 ± 3	107	−46 ± 42	0.030
Average (range), all four products ^c	1,102 (1,078–1,128)	(102–107)		
Average (range), all without bias correction	972 (823–1,151)	(78–109)		
Alpine areas, average (range) ^d	484 (304–639)	(152–320)		
GlobSnow v3.0 based regional trends	Latitude–longitude grid (°)	SWE trend with 95% confidence interval (mm decade⁻¹)	P-value of trend line	
A. Europe (Baltic)	50° N to 65° N, 15° E to 35° E	−6.3 ± 5.3	0.020	
B. Siberia (western-central)	55° N to 70° N, 65° E to 115° E	3.1 ± 2.9	0.038	
C. East Siberia	65° N to 73° N, 145° E to 175° E	8.6 ± 4.9	0.001	
D. Prairie (USA, Canada)	40° N to 50° N, 110° W to 95° W	2.8 ± 2.7	0.039	
E. Hudson Bay area	50° N to 65° N, 100° W to 70° W	−8.5 ± 3.9	0.000	

^aError bounds for snow mass estimates are obtained from spatially weighted standard deviations (calculated from spatial error variances determined from the bias-correction field).

^bThe average SWE of a certain year is defined here as the average of all grid cells that are snow covered.

^cBias-corrected GlobSnow v3.0 and bias-corrected reanalyses products: MERRA2, Crocus v7 and Brown (the number of snow courses applied for bias correction is more than 2,500 for each product).

^dAlpine areas (3 × 10⁶ km² in Eurasia and 2 × 10⁶ km² in North America) are estimated without bias correction from reanalysis products (MERRA2, Crocus v7 and Brown).

^eTrend not statistically significant (that is, $P > 0.05$, less than 95% probability for a trend different from zero).

CDR (released in 2019) with extensive ground-based snow course SWE measurements, in order to provide improved large-scale snow mass estimates with quantified uncertainty information (error bounds; see Methods for details). We apply the same bias-correction analysis to three gridded hemispheric reanalysis or reanalysis data driven snow products (Crocus v7, MERRA2 and Brown; described in the Methods). We analyse the bias by comparing SWE estimates with independent multiyear snow course observations of SWE, distributed across Eurasia and North America. The key result is the estimated absolute value of hemispheric- and continental-scale snow mass obtained for the period 1980–2018 from different snow data records. Further, an essential result is a reliable estimation of snow mass trends based on the use of the snow course dataset for validation.

The GlobSnow analysis excludes alpine areas with high topographic variability within the satellite data footprint²⁸. The bias-correction approach cannot be applied for these mountain areas owing to the distribution of hemispheric snow courses required for the bias correction (Extended Data Fig. 2) and the high degree of sub-grid variability in snow mass in areas of complex terrain. However, we provide results from the other three products—without applying bias correction—for alpine areas, although it is almost certain that these estimates are biased low¹⁸. Alpine snow is of particular importance for water resources in certain drainage basins of North America and Eurasia. We estimate that approximately 15% of the total seasonal snow area above the 40° N is in alpine regions for North America, and about 13% for Eurasia.

The absolute value of seasonal snow mass across the Northern Hemisphere is presented in Figs. 1, 2 and Table 1. Our estimate of the average

hemispheric non-alpine March snow mass is 3,062 Gt for the period 1980–2018, based on the bias-corrected GlobSnow v3.0 (corresponding to a mean SWE of 97 mm across the snow-covered area). The range provided by all investigated products after applying bias correction is 2,846–3,062 Gt (mean SWE range 90–97 mm), whereas the range from all four snow data records without bias correction is 2,433–3,380 Gt. This suggests a reduction in the uncertainty of the snow mass estimation of a factor of 4.4 for non-alpine areas above the 40° N—that is, uncertainty reduces from roughly 33% to roughly 7.4% (947 Gt to 216 Gt). Approximately 62% of non-alpine snow mass is located in Eurasia, and 38% in North America. The maximum snow mass occurs during March for the period 1980–2018, with the exception of the year 2000 in North America (Extended Data Table 1). To constrain the range in absolute values, the applied bias correction proves essential.

Earlier estimates⁸ of hemispheric maximum snow mass including alpine areas suggest values ranging from 2,500 Gt to 4,200 Gt. Combining bias-corrected minimum and maximum estimates for non-alpine areas with minimum and maximum estimates for alpine areas (Table 1), we obtain the corresponding total hemispheric snow mass, which ranges from 3,328 Gt to 3,999 Gt. The estimates of alpine snow mass are obtained from MERRA2, Crocus v7 and Brown without applying the bias correction, as the method developed here cannot be used for mountain regions with high topographic variability. Thus, the alpine snow mass estimates listed in Table 1 are probably underestimated¹⁷. Further reduction of uncertainty in hemispheric snow mass requires better-resolved estimates in alpine/mountain areas, as these areas cannot be reliably mapped or bias corrected at coarse spatial resolution

with the approach implemented here. Table 1 suggests that alpine areas have about 22–37% of the total snow mass in North America, and 9–14% in Eurasia.

Snow mass trends and patterns

The GlobSnow v3.0 dataset enables the reliable estimation of snow mass (SWE) trends. Our analysis shows that the maximum continental-scale snow mass for Eurasia is not declining, but instead showing, on average, consistent values even though regional trends are strong. The results indicate a decreasing (statistically significant, $P=0.048$) trend of -49 Gt per decade for the Northern Hemisphere March snow mass (the month of maximum snow mass), driven by a statistically significant trend ($P=0.030$) of -46 Gt per decade for North America (see Fig. 1 and Table 1). A possible reason for this behaviour is a difference in winter precipitation trends for the regions of seasonal snow cover¹⁰. A particular feature in North America is an increase in snow mass for the early years of the time series, followed by a decreasing trend since around 1990.

Unlike the other datasets, GlobSnow v3.0 does not exhibit a significant trend in systematic error over the full data record. That is, the mean difference between time series of SWE estimates and snow course observations is statistically consistent. This indicates that it is uniquely positioned to serve as the core dataset for trend determination (see Methods and Extended Data Fig. 3).

In addition to hemispheric- and global-scale estimates and trends of seasonal snow mass, our analysis shows distinct spatial patterns of snow mass distribution and its trend for 1980–2018. Trends shown in Fig. 2 indicate that snow mass has increased across large portions of Siberia and in some coastal regions (Arctic Ocean and Japan Sea coasts). Loss of snow mass is evident in Europe, in large portions of Yukon/Alaska, and in regions around Hudson Bay. Table 1 summarizes regional trends for five areas of interest. In general, the linear trends indicate large decreases in snow mass in Europe and east and west of Hudson Bay (regions A and E) during the period 1980–2018. However, in East Siberia (region C), the evolution of March snow mass during the recent years shows a dramatic increase beyond the linear trend given in Table 1 (see the lower panel of Fig. 3). Measurements by 23 weather stations in region C show a major increase in early winter solid precipitation for the two last winters of the time series. The observed sum of solid precipitation in October and November is on average 41 mm (standard deviation 8 mm), but as high as 75 mm and 63 mm for the last two winters. The high amount of early winter solid precipitation coincides with an absence of sea ice on the East Siberian Sea adjacent to region C (sea ice index, National Snow and Ice Data Center).

In contrast with the East Siberian region C, the trend in Eurasian region A (Baltic area) shows a clear gradual decline in snow mass throughout the time period 1980–2018. However, the interannual variability is very high as the region is close to the continental snow line, with typical wintertime air temperatures close to zero. In particular, the analysis of weather-station temperature data from southern Finland shows that below-average levels of snow mass coincide with above-average winter season air temperatures (which correspond to March mean temperatures close to zero).

The limitation of using only snow course data to assess snow mass distribution and trends is seen in the upper panel of Fig. 3. Altogether, nine snow courses reported SWE in region C (East Siberia) for the period 1980–2016. However, these observations are not consistent through time, as only a subset of the snow courses report in a single year. The high range of variability in SWE across snow courses from a single year emphasizes the problems of assessing trends arising from this inconsistency. Hence, the yearly GlobSnow estimates of regional SWE (Fig. 3, lower panel) show values that are slightly different to the annual March averages of the reporting snow courses. Nevertheless, the 39-year-period average levels of snow course observations and

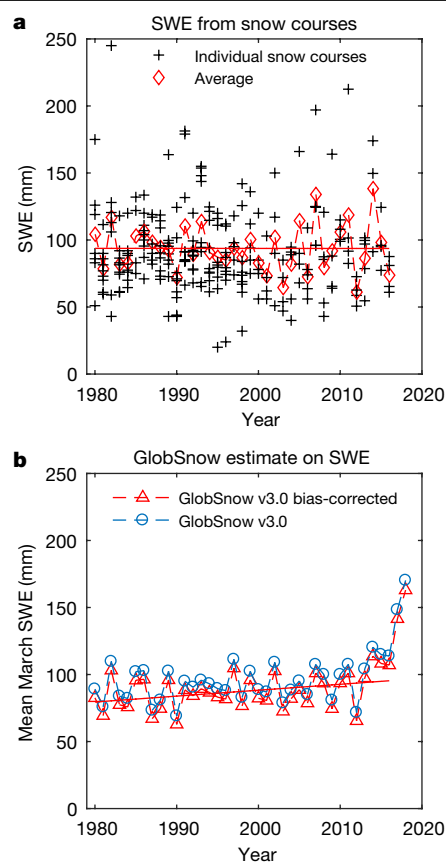


Fig. 3 | Evolution of March SWE in East Siberia. a, March observations from up to nine snow courses in region C (see Fig. 2) for 1980–2016. **b**, GlobSnow v3.0 estimates of the regional average of March SWE (1980–2018).

GlobSnow estimates on SWE correspond well to each other. Fig. 3 also illustrates regression lines fitted to both sets of data for the years 1980–2016, showing an increasing trend in GlobSnow v3.0 estimates. Regression lines do not extend to the years 2017 and 2018, because snow course observations are not yet available. As discussed above, the indicated strong increase in SWE for March 2017 and 2018 is supported by the observations of weather stations reporting the snow depth and precipitation in the area.

Summary

Here we have assessed the annual hemispheric maximum snow mass above 40° N by combining satellite data and in situ observations (Table 1). Our analysis narrows the range identified in previous snow mass evaluations (based on modelling, reanalysis and observational data analysis) by incorporating snow depth observations in the SWE retrieval, and by applying SWE snow course measurements for bias correction. Thus, our results facilitate the quantification of the snow component in Earth's water cycle, at hemispheric to regional scales. Likewise, using the results to generate an improved long-term CDR of snow mass evolution facilitates the characterization and monitoring of an essential aspect of Earth's climate system. Additionally, our results indicate that continental-scale trends over North America and Eurasia show different behaviours (Table 1). For North America the overall trend is negative, and for Eurasia the trend is negligible for the investigated non-alpine regions (Fig. 1 and Table 1). On smaller regional scales, we have identified areas with either a decrease or an increase in snow mass on both continents (Fig. 2). In most cases, the gradual and highly variable change in the amount of snow occurs throughout the available time series of 39 years (1980–2018). However, we found a

drastic increase in snow mass taking place during the last two years of the time series in the Arctic Ocean coastal region of East Siberia, which causes the overall observed regional trend (Fig. 3 and Table 1), and is also suggested by multisource data analyses of snow depth anomalies³⁰. We conclude that our trend analysis provides key information for evaluating the impacts and feedbacks of cold-season changes across mid- and high-latitude regions.

Online content

Any methods, additional references, Nature Research reporting summaries, source data, extended data, supplementary information, acknowledgements, peer review information; details of author contributions and competing interests; and statements of data and code availability are available at <https://doi.org/10.1038/s41586-020-2258-0>.

- Brown, R. D. & Mote, P. W. The response of Northern Hemisphere snow cover to a changing climate. *J. Clim.* **22**, 2124–2145 (2009).
- Derkson, C. & Brown, R. Spring snow cover extent reductions in the 2008–2012 period exceeding climate model projections. *Geophys. Res. Lett.* **39**, L19504 (2012).
- Hori, M. et al. A 38-year (1978–2015) Northern Hemisphere daily snow cover extent product derived using consistent objective criteria from satellite-borne optical sensors. *Remote Sens. Environ.* **191**, 402–418 (2017).
- Barnett, T. P., Adam, J. C. & Lettenmaier, D. P. Potential impacts of a warming climate on water availability in snow-dominated regions. *Nature* **438**, 303–309 (2005).
- IPCC. *Climate Change 2013: The Physical Science Basis* (eds Stocker, T. F. et al.) (Cambridge Univ. Press, 2013).
- Sturm, M., Goldstein, M. A. & Parr, C. Water and life from snow: a trillion dollar science question. *Wat. Resour. Res.* **53**, 3534–3544 (2017).
- Liu, J., Li, Z., Huang, L. & Tian, B. Hemispheric scale comparison of monthly passive microwave snow water equivalent products. *J. Appl. Remote Sens.* **8**, 084688 (2014).
- Mudryk, L. R., Derksen, C., Kushner, P. J. & Brown, R. Characterization of Northern Hemisphere snow water equivalent datasets, 1981–2010. *J. Clim.* **28**, 8037–8051 (2015).
- Bormann, K. J., Brown, R. D., Derksen, C. & Painter, T. H. Estimating snow-cover trends from space. *Nat. Clim. Chang.* **8**, 924–928 (2018).
- Henderson, G. R., Peings, Y., Furtado, J. C. & Kushner, P. J. Snow-atmosphere coupling in the Northern Hemisphere. *Nat. Clim. Chang.* **8**, 954–963 (2018).
- Bintanja, R. & Andry, O. Towards a rain dominated Arctic. *Nat. Clim. Chang.* **7**, 263–267 (2017).
- Flanner, M. G., Shell, K. M., Barlage, M., Perovich, D. K. & Tschudi, M. A. Radiative forcing and albedo feedback from the Northern Hemisphere cryosphere between 1979 and 2008. *Nat. Geosci.* **4**, 151–155 (2011).
- Bokhorst, S. et al. Changing Arctic snow cover: a review of recent developments and assessment of future needs for observations, modelling and impacts. *Ambio* **45**, 516–537 (2016).
- Natali, S. et al. Large loss of CO₂ in winter observed across the northern permafrost region. *Nat. Clim. Chang.* **9**, 852–857 (2019); correction **9**, 1005 (2019).
- Kittler, F. et al. Long-term drainage reduces CO₂ uptake and CH₄ emissions in a Siberian permafrost ecosystem. *Glob. Biogeochem. Cycles* **31**, 1704–1717 (2017).
- Pulliainen, J. et al. Early snowmelt significantly enhances boreal springtime carbon uptake. *Proc. Natl Acad. Sci. USA* **114**, 11081–11086 (2017).
- Wrzesien, M. et al. A new estimate of North American mountain snow accumulation from regional climate model simulations. *Geophys. Res. Lett.* **45**, 1423–1432 (2018).
- de Rosnay, P., Balsamo, G., Albergel, C., Muñoz-Sabater, J. & Isaksen, I. Initialisation of land surface variables for numerical weather prediction. *Surv. Geophys.* **35**, 607–621 (2014).
- Jeong, D., Sushama, L. & Khaliq, M. Attribution of spring snow water equivalent (SWE) changes over the northern hemisphere to anthropogenic effects. *Clim. Dyn.* **48**, 3645–3658 (2017).
- Frei, A. et al. A review of global satellite-derived snow products. *Adv. Space Res.* **50**, 1007–1029 (2012).
- Chang, A., Foster, J. & Hall, D. Nimbus-7 SMMR derived global snow cover parameters. *Ann. Glaciol.* **9**, 39–44 (1987).
- Kelly, R., Chang, A., Tsang, L. & Foster, J. A prototype AMSR-E global snow area and snow depth algorithm. *IEEE Trans. Geosci. Remote Sens.* **41**, 230–242 (2003).
- Derkson, C., Walker, A. & Goodison, B. Evaluation of passive microwave snow water equivalent retrievals across the boreal forest/tundra transition of western Canada. *Remote Sens. Environ.* **96**, 315–327 (2005).
- Pulliainen, J. Mapping of snow water equivalent and snow depth in boreal and sub-arctic zones by assimilating space-borne microwave radiometer data and ground-based observations. *Remote Sens. Environ.* **101**, 257–269 (2006).
- Derkson, C. et al. Northwest territories and Nunavut snow characteristics from a subarctic traverse: implications for passive microwave remote sensing. *J. Hydrometeorol.* **10**, 448–463 (2009).
- Lemmetyinen, J. et al. A comparison of airborne microwave brightness temperatures and snowpack properties across the boreal forests of Finland and western Canada. *IEEE Trans. Geosci. Remote Sens.* **47**, 965–978 (2009).
- Kontu, A., Lemmetyinen, J., Vehviläinen, J., Leppänen, L. & Pulliainen, J. Coupling SNOWPACK-modeled grain size parameters with the HUT snow emission model. *Remote Sens. Environ.* **194**, 33–47 (2017).
- Takala, M. et al. Estimating northern hemisphere snow water equivalent for climate research through assimilation of space-borne radiometer data and ground-based measurements. *Remote Sens. Environ.* **115**, 3517–3529 (2011).
- Brun, E. et al. Simulation of northern Eurasian local snow depth, mass, and density using a detailed snowpack model and meteorological reanalyses. *J. Hydrometeorol.* **14**, 203–219 (2013).
- Mudryk, L. et al. Terrestrial Snow Cover. *Arctic Report Card* <https://www.arctic.noaa.gov/Report-Card> (2018).

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

© The Author(s), under exclusive licence to Springer Nature Limited 2020

Methods

GlobSnow CDRs on snow extent and SWE were introduced in 2011, and can be found at <https://globsnow.info>²⁸. These products were further improved with the release of new GlobSnow snow extent and SWE product versions during 2014–2019 (ref. ³¹). The GlobSnow SWE daily time series starts from 1979 and extends to the present. The SWE values are provided in a fixed equal-area grid (EASE grid) of 625 km², that is, nominally 25 km by 25 km. Additionally, per-grid-cell estimates of the total product error are given.

The GlobSnow SWE retrieval methodology is based on Bayesian statistical inversion theory³². This method was adopted for Earth surface geophysical parameter retrieval by the inversion of a forward model of spaceborne microwave radiometer observations³³, introducing a modelling approach of spaceborne-observed scene-brightness temperature with the consideration of atmospheric effects by a statistical model. A consideration of microwave propagation in snow and the emission of snowpack in the forward model, as well as empirical modelling of forest canopy effects³⁴, were incorporated into scene-brightness temperature modelling in later studies³⁵. On the basis of these investigations, the assimilation approach²⁴—that is, a dynamically constrained statistical inversion combining brightness temperatures and weather-station observations of snow depth—was introduced and demonstrated²⁸ for hemispheric application using satellite observations of T_b at 19 GHz and 37 GHz. In particular, the method estimates the spatial and temporal variability of snow microstructure (including the effect of features such as ice crusts), described by an effective snow grain size, in order to obtain improved performance in radiometer-based retrievals of SWE. The method inherently decreases the weight of satellite data in the case of wet snow.

Compared with earlier GlobSnow versions²⁸, major changes have been made for v3.0, the version used here. These changes include spatiotemporal homogenization of applied synoptic weather-station snow depth observations, consideration of lake ice in the forward model of spaceborne-observed microwave brightness temperature³⁶, and revised modelling of forest canopy brightness temperature and microwave attenuation in the forest canopy³⁷. In order to constrain the number of variables in the SWE retrieval procedure, the snowpack is considered as a single layer with a constant density (0.24 g cm⁻³).

In addition to GlobSnow v3.0 CDR, we used three other gridded, hemispheric SWE data products to investigate the magnitude of March snow mass: Crocus v7 (ref. ³⁸), MERRA2 (refs. ^{39,40}) and Brown⁴¹. Brown uses a simple snow scheme⁴¹ driven by ERA-Interim meteorology. The MERRA2 product is the SWE produced within the MERRA2 reanalysis system; Crocus is a physical snow model driven by ERA-Interim meteorology. Product-specific spatial bias correction performed using the same methodology was applied to all four datasets in order to assess the range of variability in hemispheric- and continental-scale snow mass estimates.

We investigated the feasibility of using snow CDRs to assess hemispheric snow mass by comparing each data record with coincident snow course observations of SWE. We constructed a database of hemispheric-scale snow course observations via datasets from the former Soviet Union/Russia (FSU), Finland and Canada^{42,43}. Snow courses are transects in which SWE is sampled manually, typically at several locations, to overcome uncertainties related to local-scale spatial variability in snow conditions and land cover (sampling at up to 100 locations along courses that have lengths varying from 0.5 km to 4 km, or 0.1 km to 0.2 km in Canada). Exact measurement practices vary by country and location. Investigations of Canadian boreal sites⁴⁴ showed that as few as 20 distributed snow samples along a snow course provide information on the landscape mean with an accuracy better than 20%. Further, our investigation of snow course reference data for Eurasia and North America indicated that a typical exponential autocorrelation length of snow course observations is from 150 km

to 250 km depending on region, indicating that snow courses provide useful information on regional-scale SWE, and so can be meaningfully compared with satellite or reanalysis estimates having a spatial resolution of about 25 km (although they are still limited by the impacts of land cover and sparse observations through time).

Concerning GlobSnow v3.0 analyses, the applied hemispheric snow course in situ reference data provided 2,687 distributed regional observations with coincident satellite retrievals, comprising all together 343,241 observations over the time period 1979–2018 (all months). The results indicate an overall root mean square error (RMSE) of 46.2 mm, and an RMSE of 31.8 mm when cases with reference SWE greater than 150 mm are excluded (12% of the reference observations are omitted). SWE values of more than 150 mm can be considered a rough threshold when microwave radiometer measurements no longer provide useful information on SWE⁴⁵. For Eurasia, the improvement when compared with earlier GlobSnow versions is 5.8% in RMSE and an increase from 0.55 to 0.60 in correlation coefficient. This performance improvement is obtained by the advances discussed above concerning the forward modelling of T_b and the homogenization of synoptic weather-station snow depth observations.

We investigated the feasibility of the four mentioned snow data records for trend analyses by analysing the monthly bias across individual years of the time series. This is performed separately for North America (Canada) and Eurasia by using more than 2,500 snow courses observed for March with coincident gridded SWE data available from each product (Extended Data Fig. 3). For Eurasia, we selected snow courses that provide observations at least for one year within three different reference periods (1980–1984, 1999–2003, 2012–2016) for the analysis. For North America, the selected snow courses include observations at least for two periods (1980–1984, 1999–2003).

In the case of GlobSnow v3.0 CDR, annual biases for March are close to zero for Eurasia. Negative bias values are observed for North America. The trends in observed annual biases are insignificant (P -values of 0.89 and 0.81 for Eurasia and Canada, respectively), suggesting that the GlobSnow v3.0 CDR is internally consistent and so applicable for trend analysis of snow mass. If these trends of annual biases (Extended Data Fig. 3) are considered across the full time series, the obtained values are well within the 95% confidence intervals given in Table 1 (analysed change in March bias: 2 Gt decade⁻¹ (0.10 mm decade⁻¹ for SWE) for Eurasia and -7 Gt decade⁻¹ (-0.63 mm decade⁻¹) for North America). The changes in bias and the associated P -values for the other snow data records indicate a systematic temporal change through time for Eurasia that is not evident with GlobSnow: MERRA2, P -value of 2.7×10^{-3} and -70 Gt decade⁻¹ (-3.3 mm decade⁻¹) change in bias; Crocus v7, 6.7×10^{-3} and -67 Gt decade⁻¹ (-3.2 mm decade⁻¹); Brown, 2.7×10^{-3} and -71 Gt decade⁻¹ (-3.4 mm decade⁻¹). Note that corresponding trends for Eurasia obtained by the three reanalysis database products show values from -34 Gt decade⁻¹ to -85 Gt decade⁻¹, indicating that these trends are influenced by the changing product biases. This highlights the uniquely consistent nature of GlobSnow v3.0 for hemispheric trend analysis.

The primary challenge in considering regional bias arises from the inconsistent distribution of reference snow courses. In order to analyse the magnitude of these uncertainties, we developed a technique to assess the spatial bias. First, we calculate a mean SWE BIAS_{*i*} (mm) relative to the reference observations at snow course i from all observations of that particular snow course. Denoting REF_{*i,t*} as the SWE reference observation for snow course i at time step t , and EST_{*i,t*} as the corresponding estimate, we calculate the bias for snow course i across the whole time series by:

$$\text{BIAS}_i = \frac{1}{N_i} \sum_{t=1}^{N_i} (\text{EST}_{i,t} - \text{REF}_{i,t}) \quad (1)$$

A fraction of 50% of hemispheric snow courses show SWE estimation errors smaller than 33.1 mm (the RMSE value of SWE estimation errors

for different years observed at a particular location when GlobSnow v3.0 March SWE data are compared with the concurrent snow course observations at the same location). Further, the median bias is close to zero, even though high underestimation of SWE is evident for a small proportion of snow courses with very deep snow (Extended Data Fig. 4). For other investigated SWE datasets, these median RMSE characteristics show slightly higher values: 37.8 mm for Brown, 40.2 mm for Crocus v7 and 48.2 mm for MERRA2.

The map of GlobSnow v3.0 spatial bias for March, according to Equation (1), is obtained by analysing the observations from 2,636 snow courses reporting for March (see Extended Data Fig. 2, which also shows the locations of snow courses). The spatial behaviour is interpolated from the snow-course-observed average bias values by ordinary kriging interpolation (and by the nearest-neighbour method for comparison). Thereafter, this constant map is applied to SWE estimates of each year in order to yield a bias-corrected estimate of SWE (mm), and, through that, a bias-corrected estimate of snow mass (Gt):

$$\text{Snow mass} = \frac{625 \times 10^3 N_{\text{GridCells}} (\langle \text{SWE} \rangle - \langle \text{BIAS} \rangle)}{10^9} \quad (2)$$

where $N_{\text{GridCells}}$ is the number of EaseGrid cells (grid for all snow data records) in the area of interest. $\langle \text{SWE} \rangle$ and $\langle \text{BIAS} \rangle$ are, respectively, the GlobSnow v3.0 provided mean SWE and spatially averaged bias values (Extended Data Fig. 3) on the area of interest. The bias correction by Equation (2) can be applied to correct the March SWE map of all years, because the linear temporal change of bias does not affect the average over the period. Moreover, in the case of GlobSnow v3.0, the average bias remains constant through the full time series (Extended Data Fig. 3). The bias correction yields higher values of SWE than the original GlobSnow v3.0 product, particularly for the Pacific regions and eastern parts of Canada (Extended Data Fig. 1 and Extended Data Fig. 2). When the bias correction is applied to all investigated hemispheric SWE datasets (GlobSnow v3.0, Crocus v7, Brown and MERRA2), the snow mass estimates converge towards similar values at both hemispheric and continental scales (Table 1 and Extended Data Table 2).

We consider three sources of error when estimating the statistical uncertainty of the snow mass estimates: first, the temporal variability of bias observed for each individual snow course expressed with a variance (Extended Data Fig. 3); second, the spatially weighted uncertainty of estimation of bias correction for the area of interest (Extended Data Fig. 2); and third, a spatially weighted leave-one-out Monte Carlo simulation analysis that assesses the effect if any single snow course is omitted when determining the bias correction. Total statistical errors for estimating mean values (Table 1) are directly obtained by kriging interpolation, which provides an estimate of the spatial variance of the bias correction for all grid cells. Thus, the \pm uncertainty values given in Table 1 and Extended Data Table 2 include the first and second uncertainty contributions noted above. Additionally, we report the maximum (worst-case) snow mass estimation errors (Extended Data Table 1) for GlobSnow v3.0 derived estimates—that is, the error levels in the mean snow mass estimates that are obtained if the snow courses most susceptible to estimation error are omitted when estimating the bias map. This worst-case error is actually a realistic measure of the possible error that may hinder the performance of the bias correction. If an additional snow course were located in a region that is poorly represented by the present snow course network, the expected maximum error would be equivalent to this determined worst-case error (note that spatially distributed multiple errors would probably have different signs, cancelling each other in summation).

We carried out spatial analysis in order to consider the spatial inhomogeneity of snow course observation networks (the spatial density of snow courses is sparse at the highest latitudes in both continents). The applied kriging interpolation does this in a statistically robust manner. Thus, we weight each snow course according to its spatial influence,

determined by the estimated spatial variance of the interpolation field. Comparison of kriging interpolation for bias correction obtained by the nearest-neighbour method shows very small differences, highlighting the robustness of the approach: the effect on hemispheric-scale snow mass estimates ranges from -61 Gt to $+71$ Gt (-2.1% to $+2.3\%$), depending on the snow data record. The maximum-error consideration also tackles the problem of inconsistently distributed snow courses.

In the trend analyses of Fig. 1 and Table 1, a simple linear regression was applied as a baseline, providing confidence intervals and P -values. For comparison, a Theil–Sen trend estimation was also carried out, as that method is less sensitive to data outliers. This analysis yielded trends of $+3.0$ Gt decade $^{-1}$, -53.3 Gt decade $^{-1}$ and -51.5 Gt decade $^{-1}$ for Eurasia, North America and the Northern Hemisphere, respectively. These trends are stronger than a simple linear regression for North America and the Northern Hemisphere as a whole, and yield a small positive trend for Eurasia (not statistically significant). When assessing trend significance, it can be necessary to prewhiten the data in order to mitigate the effects of lag-1 autoregression⁴⁶. However, the trends analysed here lack substantial lag-1 autocorrelations, and so we did not carry out prewhitening.

Data availability

The data are available from ref. 47. The same dataset is also available from the GlobSnow service: http://www.globsnow.info/swe/archive_v3.0/; http://www.globsnow.info/swe/archive_v3.0/L3A_daily_SWE/ (daily data); http://www.globsnow.info/swe/archive_v3.0/L3B_monthly_SWE/ (monthly data); http://www.globsnow.info/swe/archive_v3.0/L3B_monthly_biascorrected_SWE/ (bias-corrected data); and http://www.globsnow.info/swe/archive_v3.0/auxiliary_data/ (all auxiliary data).

Code availability

Codes are available from <https://github.com/fmidev/GlobSnow3.0> and http://www.globsnow.info/swe/archive_v3.0/source_codes/.

- Metsämäki, S. et al. Introduction to GlobSnow Snow Extent products with considerations for accuracy assessment. *Remote Sens. Environ.* **156**, 96–108 (2015).
- Lehtinen, M. *On Statistical Inversion Theory. Theory and Applications of Inverse Problems* (Longman, 1988).
- Pulliainen, J., Kärnä, J.-P. & Hallikainen, M. Development of geophysical retrieval algorithms for the MIMR. *IEEE Trans. Geosci. Remote Sens.* **31**, 268–277 (1993).
- Kruopis, N. et al. Passive microwave measurements of snow-covered forest areas in EMAC'95. *IEEE Trans. Geosci. Remote Sens.* **37**, 2699–2705 (1999).
- Pulliainen, J., Grandell, J. & Hallikainen, M. HUT snow emission model and its applicability to snow water equivalent retrieval. *IEEE Trans. Geosci. Remote Sens.* **37**, 1378–1390 (1999).
- Lemmetyinen, J. et al. Correcting for the influence of frozen lakes in satellite microwave radiometer observations through application of a microwave emission model. *Remote Sens. Environ.* **115**, 3695–3706 (2011).
- Cohen, J. et al. The effect of boreal forest canopy in satellite snow mapping—a multisensor analysis. *IEEE Trans. Geosci. Remote Sens.* **53**, 6593–6607 (2015).
- Brun, E. et al. Simulation of northern Eurasian local snow depths, mass and density using a detailed snowpack model and meteorological reanalysis. *J. Hydrometeorol.* **14**, 203–219 (2013).
- Gelaro, R. et al. The Modern-Era Retrospective Analysis for Research and Applications, version 2 (MERRA-2). *J. Clim.* **30**, 5419–5454 (2017).
- Global Modeling and Assimilation Office. MERRA-2 tagv1_2d_Ind_Ndx: 2d, 1-hourly, time-averaged, single-level, assimilation, land surface diagnostics V5.12.4, Greenbelt, MD, USA; <https://doi.org/10.5067/RKPH78KCIYIT> (Goddard Earth Sciences Data and Information Services Center, 2015).
- Brown, R., Brasnett, B. & Robinson, D. Gridded North American monthly snow depth and snow water equivalent for GCM evaluation. *Atmos.-Ocean* **41**, 1–14 (2003).
- Brönnimann, S. et al. Observations for reanalyses. *Bull. Am. Meteorol. Soc.* **99**, 1851–1866 (2018).
- Brown, R. D., Fang, B. & Mudryk, L. Update of Canadian historical snow survey data and analysis of snow water equivalent trends, 1967–2016. *Atmos.-Ocean* **59**, 149–156 (2019).
- Neumann, N., Derksen, C., Smith, C. & Goodison, B. Characterizing local scale snow cover using point measurements during winter season. *Atmos.-Ocean* **44**, 257–269 (2006).
- Mätzler, C. Passive microwave signatures of landscapes in winter. *Meteorol. Atmos. Phys.* **54**, 241–260 (1994).
- Yue, S., Pilon, P., Phinney, B. & Cavadias, G. The influence of autocorrelation on the ability to detect trend in hydrological series. *Hydrol. Processes* **16**, 1807–1829 (2002).
- Luojus, K., Pulliainen, J., Takala, M., Lemmetyinen, J. & Moisander, M. GlobSnow v3.0 snow water equivalent (SWE). *Pangaea* <https://doi.org/10.1594/PANGAEA.911944> (2020).

Acknowledgements This work is supported by the European Space Agency (ESA) Snow Climate Change Initiative (CCI) project (grant 4000124098/18/I-NB); the ESA SnowPEX project (4000111278/14/I-LG); the Academy of Finland Centre of Excellence (118780); the Academy of Finland ASTRA-Snow project (325397); and the Nordic Centre of Excellence Climate Change Effects on the Epidemiology of Infectious Diseases and the Impacts on Northern Societies (CLINF; 76413).

Author contributions J.P. initiated the work and made a major contribution towards developing the methodology and writing the manuscript. K.L., C.D., L.M. and J.L. contributed to data analysis and writing of the manuscript, and supported the conclusions. M.S.

contributed to the writing and editing/revising of the manuscript. J.I., M.T., J.C., T.S. and J.N. contributed to the analysis and processing of satellite and in situ data.

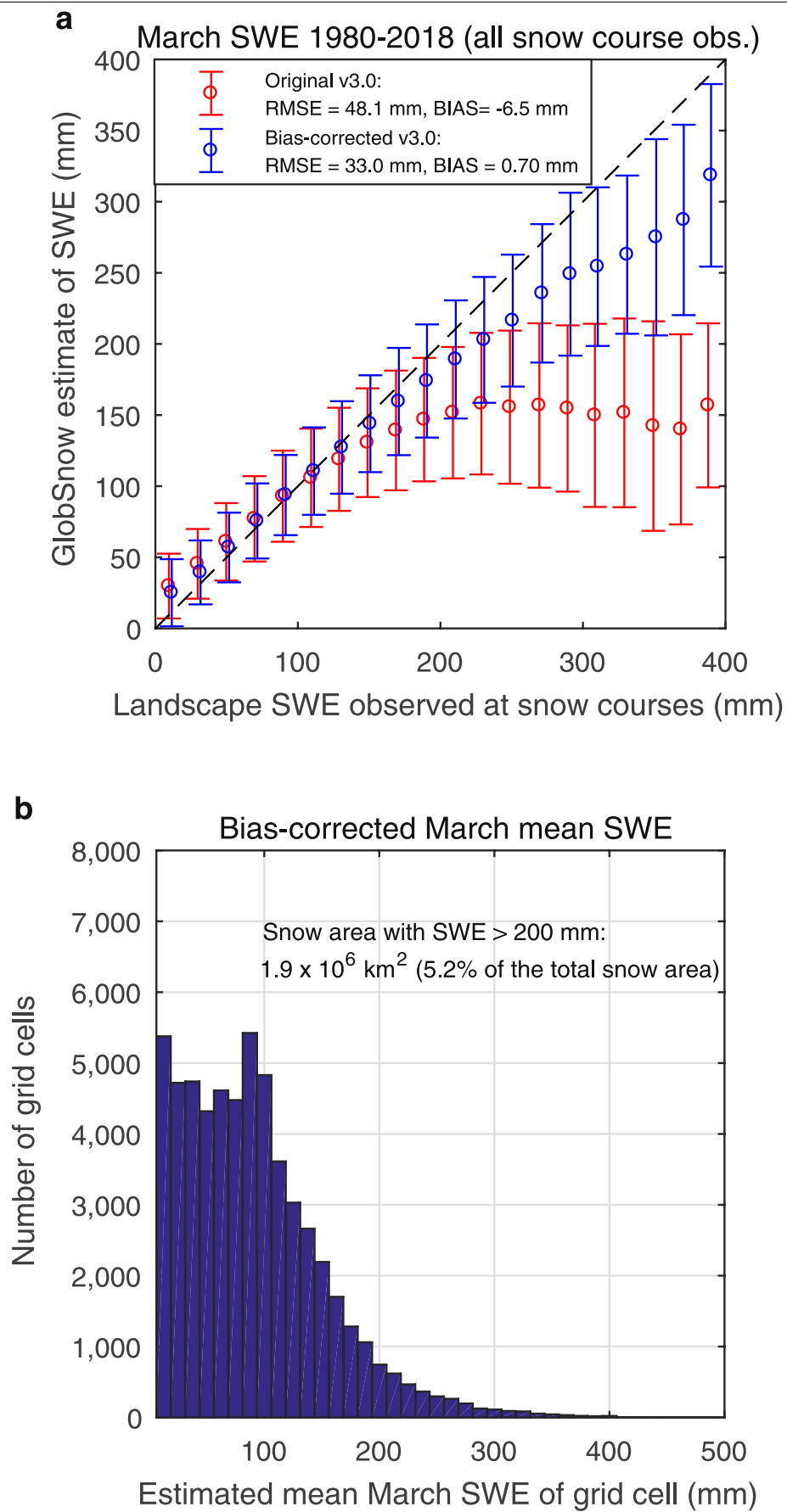
Competing interests The authors declare no competing interests.

Additional information

Correspondence and requests for materials should be addressed to J.P.

Peer review information *Nature* thanks Fanny Larue and the other, anonymous, reviewer(s) for their contribution to the peer review of this work.

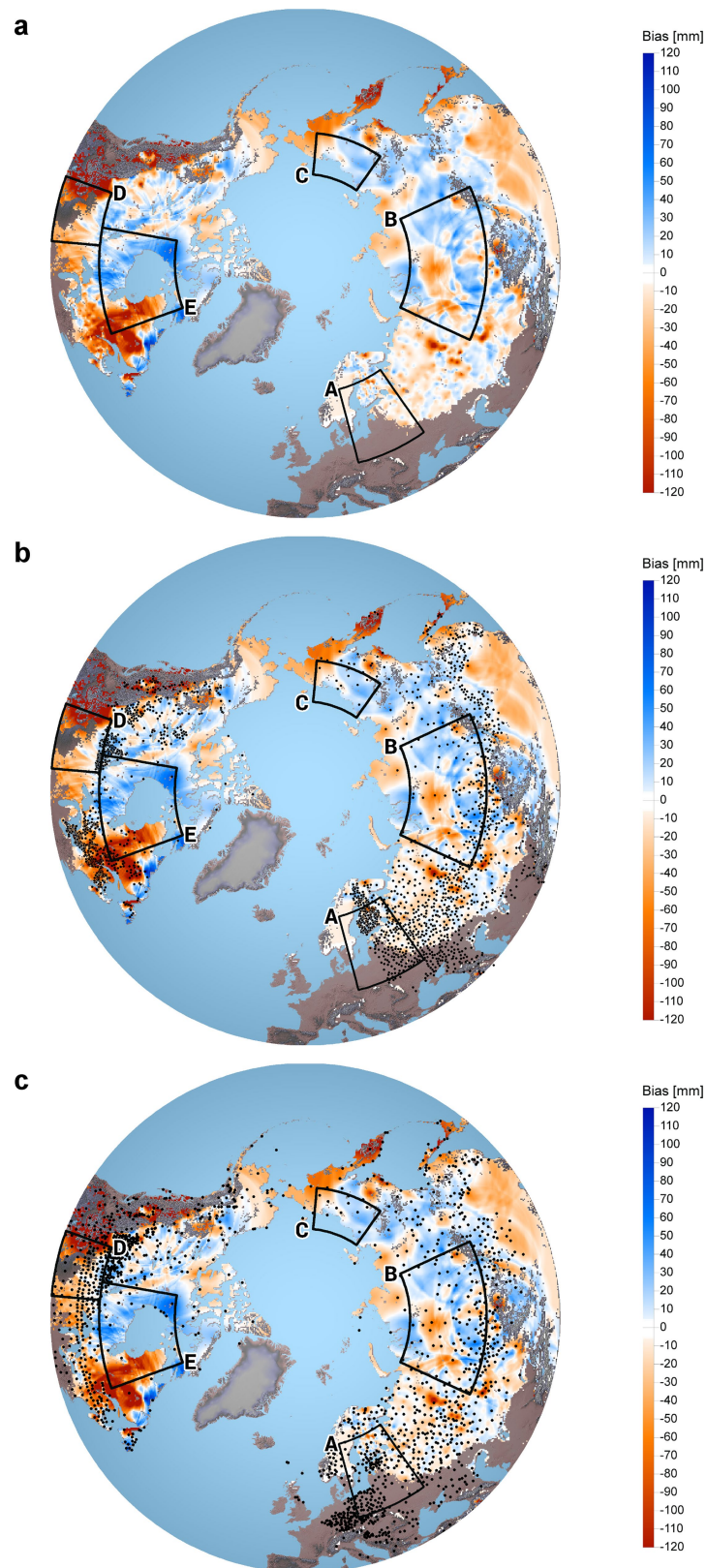
Reprints and permissions information is available at <http://www.nature.com/reprints>.



Extended Data Fig. 1 | See next page for caption.

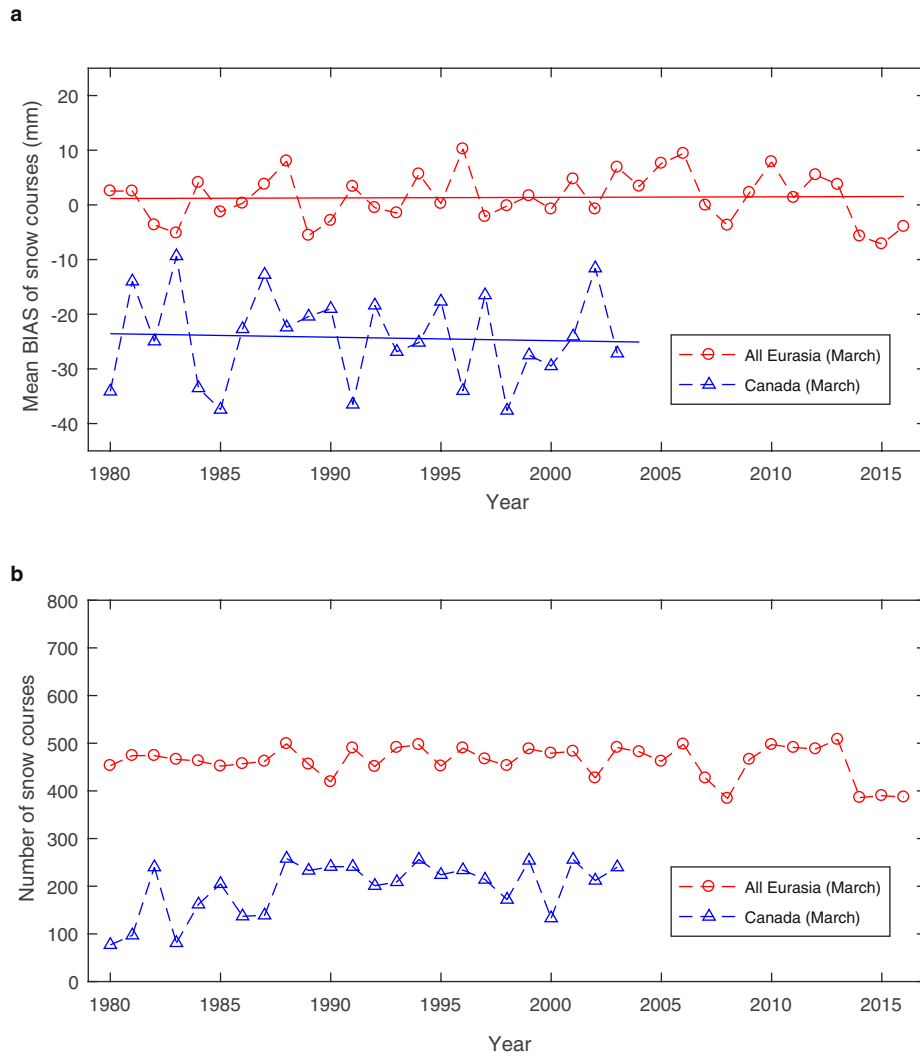
Extended Data Fig. 1 | Hemispheric SWE retrievals for March. a, Scatterplot of GlobSnow v3.0 SWE estimates versus interval-stratified in situ (landscape) SWE data for all snow course observations, with \pm standard deviations. As the original GlobSnow approach is based on microwave radiometry, it tends to underestimate SWE with high levels of SWE (more than 150 mm) owing to

saturation of the microwave signal. The bias-correction approach mitigates this problem. **b,** Histogram showing bias-corrected GlobSnow v3.0 estimates of mean March SWE (x axis) across the period 1980–2018 for all grid cells with mean SWE values of more than 0 mm.



Extended Data Fig. 2 | Map showing the spatial distribution of the SWE estimation bias. a, Kriging-interpolated map for March calculated from biases observed at the locations of 2,636 snow courses. **b,** The same map, but also

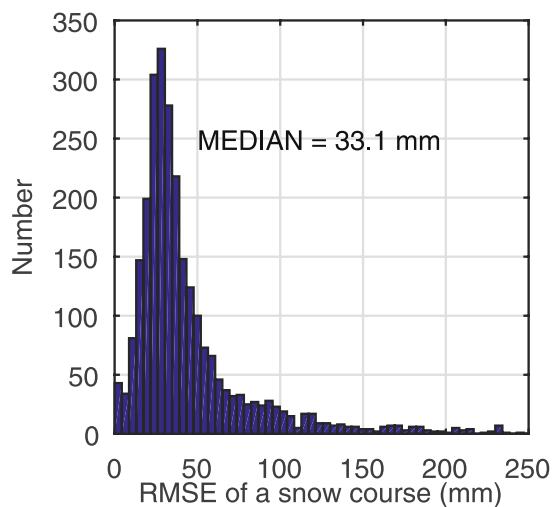
indicating the locations of snow courses (black dots). **c,** Weather stations that report snow depth (black dots). A–E are the dedicated areas that we use to investigate regional trends.



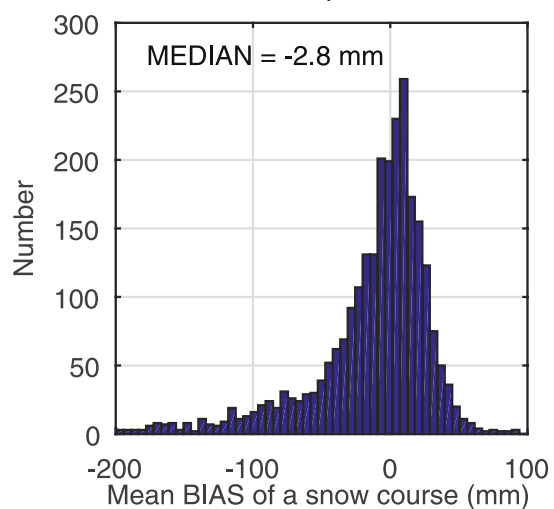
Extended Data Fig. 3 | Evolution of the annual bias in GlobSnow SWE estimates for March. In other words, the figure shows the evolution of the systematic SWE estimation error for the period of maximum snow mass in the Northern Hemisphere. **a**, We calculated bias for the March observations of a given year, separately for each snow course (locations shown in Extended Data Fig. 2). We then averaged these snow-course-stratified biases over both continents, with results shown here. The P -values for trend lines are 0.89 and

0.81 for Eurasia and North America (Canada), respectively, indicating negligible trends. The other assessed SWE datasets are not directly applicable to the trend analysis in Eurasia, as the bias compared with snow course observations changes systematically with time (P -values of bias trend lines in Eurasia are 6.7×10^{-3} for Crocus v7, 2.7×10^{-3} for Brown and 2.7×10^{-3} for MERRA2). **b**, About 400 snow courses for Eurasia and 200 for North America provide observations throughout the investigated time period.

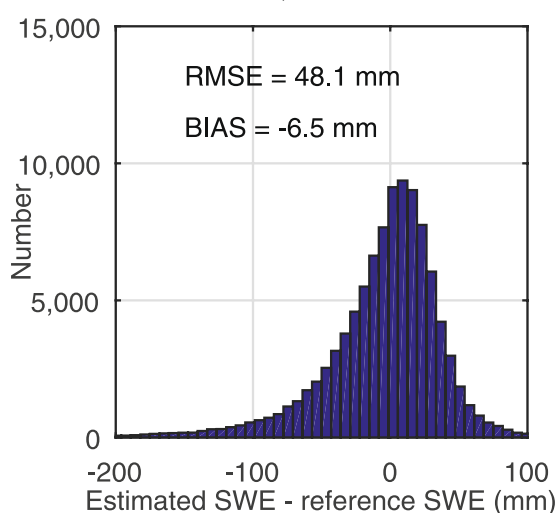
a RMSE for 2,636 hemispheric snow courses



b BIAS for 2,636 hemispheric snow courses



c Residuals for 100,651 March observations



Extended Data Fig. 4 | Histograms showing hemispheric SWE retrieval accuracy for March. a. RMSE. **b.** Bias determined for each of 2,636 snow courses. **c.** Residual errors for all 100,651 observations from March throughout the GlobSnow v3.0 time series.

Extended Data Table 1 | Monthly GlobSnow v3.0 estimates of snow mass

Region above LAT 40°, month of year	Area of seasonal snow [10 ⁶ km ²] (min/ave/max)	Mean monthly snow mass with error bounds 1980-2018 (Gt) ^a	Snow mass trend and its 95% confidence interval (Gt/10y)	p-value of trend line
Northern Hemisphere				
February	29/33/35	2,655 ± 36 (± 5)	-57 ± 42	0.010
March	29/32/35	3,062 ± 35 (± 5)	-49 ± 49	0.048
April	24/28/32	2,543 ± 95 (± 6)	-49 ± 53 ^b	0.073
May ^c	NA/12/22	1,571 ± 92 (± 8)	-80 ± 62	0.014
Eurasia				
February	19/22/24	1,720 ± 26 (± 4)	-5 ± 23 ^b	0.643
March	19/21/23	1,934 ± 35 (± 4)	-3 ± 25 ^b	0.809
April	15/18/24	1,613 ± 95 (± 4)	-19 ± 41 ^b	0.361
May ^c	NA/7/14	974 ± 92 (± 5)	-62 ± 56	0.031
North America				
February	9/11/12	935 ± 36 (± 3)	-51 ± 32	0.002
March	9/11/12	1,128 ± 31 (± 3)	-46 ± 42	0.030
April	7/9/11	930 ± 38 (± 4)	-30 ± 34 ^b	0.081
May ^c	NA/5/8	597 ± 19 (± 6)	-18 ± 22 ^b	0.104

^aError bounds for snow mass estimates are given by spatially weighted estimated maximum errors and, in parenthesis, by spatially weighted standard deviations (standard deviation of the mean mass estimate).

^bNot statistically significant.

^cBeginning of May only (one-week period around May 7). NA, not available.

Extended Data Table 2 | March snow mass from various data sources

Non-alpine regions above LAT 40°	Area of seasonal snow (10 ⁶ km ²) (min/ave/max) ^a	Bias-corrected estimate with error bounds ^b (Gt)	Original estimate (Gt)	Difference between estimates (Gt)
Northern Hemisphere	29/32/35			
GlobSnow v3.0		3,062 ± 5	2,737	-330 (-10.8%)
MERRA2		2,898 ± 7	3,380	482 (16.6%)
Crocus v7		2,946 ± 7	2,916	-30 (-1.0%)
Brown		2,846 ± 7	2,433	-413 (-14.5%)
Average (range) of all		2,938 (2,846-3,062)	2,867 (2,433-3,380)	-71 (-2.4%)
Eurasia	19/21/23			
GlobSnow v3.0		1,934 ± 4	1,831	-81 (-4.2%)
MERRA2		1,807 ± 5	2,229	422 (23.4%)
Crocus v7		1,868 ± 5	1,910	42 (2.2%)
Brown		1,736 ± 4	1,610	-126 (-7.3%)
Average (range) of all		1,836 (1736-1934)	1,895 (1,610-2,229)	59 (3.2%)
North America	9/11/12			
GlobSnow v3.0		1,128 ± 3	906	-222 (-19.7%)
MERRA2		1,091 ± 4	1,151	60 (5.5%)
Crocus v7		1,078 ± 4	1,006	-72 (-6.7%)
Brown		1,110 ± 5	823	-287 (-25.9%)
Average (range) of all		1,102 (1,078-1,128)	972 (823-1151)	-130 (-11.8%)
Alpine regions above LAT 40°	Area of alpine snow (10 ⁶ km ²)		Original estimate (Gt)	Average of all three products (Gt)
Northern Hemisphere	5			
MERRA2			709	709
Crocus v7			937	
Brown			482	
Eurasia	3			
MERRA2			200	225
Crocus v7			298	
Brown			178	
North America	2			
MERRA2			509	484
Crocus v7			639	
Brown			304	

^aThe same snow area masks were applied to each data product, excluding alpine regions for all products.
^bError bounds for snow mass estimates are obtained from spatially weighted standard deviations (calculated from spatial error variances determined from the bias-correction field).

Initial Upper Palaeolithic *Homo sapiens* from Bacho Kiro Cave, Bulgaria

<https://doi.org/10.1038/s41586-020-2259-z>

Received: 30 July 2019

Accepted: 24 February 2020

Published online: 11 May 2020

 Check for updates

Jean-Jacques Hublin^{1,2✉}, Nikolay Sirakov³, Vera Aldeias⁴, Shara Bailey^{1,5}, Edouard Bard⁶, Vincent Delvigne^{7,8}, Elena Endarova⁹, Yoann Fagault⁶, Helen Fewlass¹, Mateja Hajdinjak¹⁰, Bernd Kromer¹, Ivaylo Krumov¹¹, João Marreiros^{4,12}, Naomi L. Martisius¹³, Lindsey Paskulin¹⁴, Virginie Sinet-Mathiot¹, Matthias Meyer¹⁰, Svante Pääbo¹⁰, Vasil Popov¹⁵, Zeljko Rezek^{1,16}, Svoboda Sirakova³, Matthew M. Skinner^{1,17}, Geoff M. Smith¹, Rosen Spasov¹⁸, Sahra Talamo^{1,19}, Thibaut Tuna⁶, Lukas Wacker²⁰, Frido Welker^{1,21}, Arndt Wilcke²², Nikolay Zahariev²³, Shannon P. McPherron¹ & Tsenka Tzanova¹

The Middle to Upper Palaeolithic transition in Europe witnessed the replacement and partial absorption of local Neanderthal populations by *Homo sapiens* populations of African origin¹. However, this process probably varied across regions and its details remain largely unknown. In particular, the duration of chronological overlap between the two groups is much debated, as are the implications of this overlap for the nature of the biological and cultural interactions between Neanderthals and *H. sapiens*. Here we report the discovery and direct dating of human remains found in association with Initial Upper Palaeolithic artefacts², from excavations at Bacho Kiro Cave (Bulgaria). Morphological analysis of a tooth and mitochondrial DNA from several hominin bone fragments, identified through proteomic screening, assign these finds to *H. sapiens* and link the expansion of Initial Upper Palaeolithic technologies with the spread of *H. sapiens* into the mid-latitudes of Eurasia before 45 thousand years ago³. The excavations yielded a wealth of bone artefacts, including pendants manufactured from cave bear teeth that are reminiscent of those later produced by the last Neanderthals of western Europe^{4–6}. These finds are consistent with models based on the arrival of multiple waves of *H. sapiens* into Europe coming into contact with declining Neanderthal populations^{7,8}.

Fragmentary specimens from the sites of Kent's Cavern (United Kingdom)⁹ and Cavallo (Italy)¹⁰ have been claimed to document the earliest presence of our species in western Europe, between 44,200–41,500 calibrated years before present (cal. BP; taken as AD 1950) for the former and between 45,000–43,000 cal. BP for the latter. However, these dates are based on the archaeological contexts of the specimens rather than direct dating, and—in both cases—the exact stratigraphic origin of the fossils is debated^{11,12}. In the absence of directly dated fossil remains, reconstructing the timing of the expansions of *H. sapiens* into Europe rests on hypotheses concerning the makers of various so-called 'transitional' artefact assemblages at the advent of the Upper Palaeolithic.

Bacho Kiro Cave is located 5 km west of Dryanovo (Bulgaria), on the northern slope of the Balkan mountain range (Stara Planina) and about 70 km south of the Danube River (Extended Data Fig. 1b). The

site formed at the mouth of a large karstic system and its deposits encompass late Middle Palaeolithic and early Upper Palaeolithic occupations. Bacho Kiro Cave was excavated by D. Garrod in 1938, but is best known from more extensive excavations (in 1971 to 1975) by a team led by B. Ginter and J. Kozłowski¹³. The excavations in the 1970s yielded fragmentary human remains¹³ that were subsequently lost. In 2015, the National Archaeological Institute with Museum in Sofia and the Department of Human Evolution at the Max Planck Institute for Evolutionary Anthropology resumed work at Bacho Kiro Cave with the goals of clarifying the chronology (which had previously been based on a handful of inconsistent radiocarbon ages¹⁴) and the biological nature of the makers of the lithic assemblages. Two sectors with similar and well-preserved sequences were re-excavated: the Main sector and the previously unexcavated Niche 1 sector, located on the south and east

¹Department of Human Evolution, Max Planck Institute for Evolutionary Anthropology, Leipzig, Germany. ²Chaire Internationale de Paléoanthropologie, Collège de France, Paris, France.

³National Institute of Archaeology with Museum, Bulgarian Academy of Sciences, Sofia, Bulgaria. ⁴Interdisciplinary Centre for Archaeology and the Evolution of Human Behaviour, Universidade do Algarve, Faro, Portugal. ⁵Department of Anthropology, New York University, New York, NY, USA. ⁶CEREGE, Aix Marseille University, CNRS, IRD, INRAE, Collège de France, Aix-en-Provence, France. ⁷Service de Préhistoire, University of Liège, Liège, Belgium. ⁸CNRS, UMR 5199 PACEA, University of Bordeaux, Pessac, France. ⁹National History Museum, Sofia, Bulgaria. ¹⁰Department of Evolutionary Genetics, Max Planck Institute for Evolutionary Anthropology, Leipzig, Germany. ¹¹History Museum, Belogradchik, Bulgaria. ¹²TraCEr, Monrepos Archaeological Research Centre and Museum for Human Behavioural Evolution, RGZM, Mainz, Germany. ¹³Department of Anthropology, University of California, Davis, Davis, CA, USA. ¹⁴Department of Archaeology, University of Aberdeen, Aberdeen, UK. ¹⁵Institute of Biodiversity and Ecosystem Research, Bulgarian Academy of Sciences, Sofia, Bulgaria. ¹⁶University of Pennsylvania Museum of Archaeology and Anthropology, Philadelphia, PA, USA. ¹⁷School of Anthropology and Conservation, University of Kent, Canterbury, UK. ¹⁸Archaeology Department, New Bulgarian University, Sofia, Bulgaria. ¹⁹Department of Chemistry 'G. Ciamician', University of Bologna, Bologna, Italy. ²⁰Department of Earth Sciences, ETH Zurich, Zurich, Switzerland. ²¹Evolutionary Genomics Section, Globe Institute, University of Copenhagen, Copenhagen, Denmark. ²²Department of Cell Therapy, Fraunhofer Institute for Cell Therapy and Immunology, Leipzig, Germany. ²³Archaeology Department, New Bulgarian University, Sofia, Bulgaria. ✉e-mail: hublin@eva.mpg.de

sides, respectively, of the excavation from the 1970s (Extended Data Fig. 1a). At the base of the sequence (Supplementary Information section 1) and overlying the bedrock, layer K has a relatively low density of Middle Palaeolithic artefacts. Sedimentologically, the contact of layer K with the overlying layer J is gradual and the artefact densities remain low; however, the upper part of layer J contains artefacts identical to those in layer I. Based on the radiocarbon dates³, layer J represents more than 3,000 years of accumulation. Layer I represents an intensification of the trends seen in layer J. Layer I is an easily recognized and archaeologically rich organic deposit that spans from 45,820 to 43,650 cal. BP (95% modelled range) and yields an assemblage that was initially described as 'Bachokirian', but is now considered a variant of the Initial Upper Palaeolithic (IUP) industry¹⁵ (Extended Data Figs. 2–4, Supplementary Information section 2). Layer I is capped by water-laid deposits (layers H and G) that have little archaeological content. More than 1.7 m of deposits, containing low densities of Upper Palaeolithic artefacts, overlies layer G in the Main sector.

We found a hominin second lower molar (specimen code F6-620) (Extended Data Fig. 5a) in the upper part of layer J. The crown dimensions of this tooth place it at the high end of both the Neanderthal and the Upper Palaeolithic *H. sapiens* range (Extended Data Table 1). With the exception of a moderately expressed—but divided—middle trigonid crest, all of the morphological trait expressions found in F6-620 align the tooth with *H. sapiens* (Supplementary Information section 3). The expression of a middle trigonid crest observed on the second lower molar from Bacho Kiro Cave is present in 10% of these teeth in some groups of humans today¹⁶ and in 8% of early *H. sapiens*¹⁷. The pulp chamber is hypotaurodont¹⁸, a condition that is common in some recent human groups¹⁹ and is unlike the hypertaurodont molars of Neanderthals²⁰. The four-cusp configuration of the second lower molar from Bacho Kiro Cave is absent in Neanderthals. Our geometric morphometric analysis of the enamel–dentine junction also clearly assigns the specimen to *H. sapiens* (Extended Data Fig. 5b).

We screened 1,271 non-identifiable bones and teeth using matrix-assisted laser desorption/ionization time-of-flight mass spectrometry (MALDI–TOF MS) collagen-peptide mass fingerprinting (also known as ZooMS²¹) to identify hominin remains, with the additional aims of providing accurate molecular identifications for radiocarbon-dated specimens and of enlarging our understanding of the species composition of the fauna. ZooMS screening identified six hominin bone fragments (Extended Data Fig. 6, Supplementary Information section 4), of which four come from layer I in Niche 1, one from layer B in the Main sector (Extended Data Fig. 1) and one from the interface of layers 6a and 7 of the excavations in the 1970s¹³. Including the F6-620 tooth, we recovered five hominin specimens in total from the IUP layers. The calibrated radiocarbon dates of the 4 ZooMS-identified human fragments range from 46,790 to 42,810 cal. BP at 95.4% probability (Fig. 1). These ages are in full agreement with the modelled boundaries of layer I (45,820–43,650 cal. BP at 95.4%), which includes the 4 humans and 21 other dates on modified fauna³. Therefore, to our knowledge, these bones represent the oldest European Upper Palaeolithic hominins recovered to date.

We extracted DNA^{22,23} from F6-620 and the six hominin bone fragments identified using ZooMS. We performed library preparation²⁴, enrichment of human mitochondrial DNA (mtDNA)²⁵ and sequencing, which enabled us to recover between 13,856 and 795,043 unique mtDNA fragments (Supplementary Information section 5). The frequencies of cytosine-to-thymine substitutions, which are characteristic of ancient-DNA base damage, ranged from 13.5% to 54.9% at the 5' ends and from 9.4% to 42.2% at the 3' ends of these fragments (Extended Data Fig. 7), which suggests that at least some of the fragments are of ancient origin. After restricting analyses to putatively deaminated DNA fragments to remove contamination by recent human DNA, sequence coverage of the mitochondrial genome enabled us to reconstruct six full mitochondrial genomes out of seven. The mtDNA sequences of F6-620

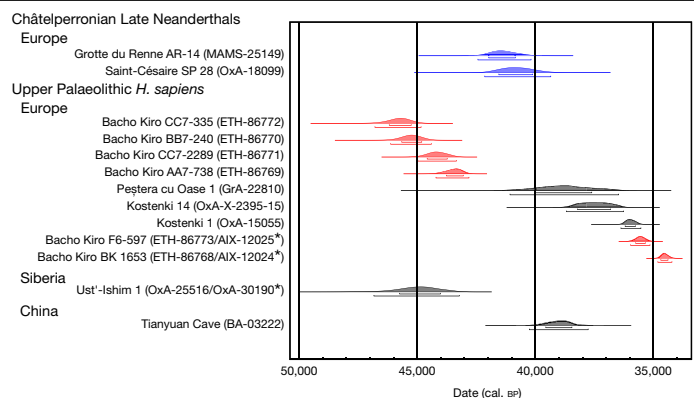


Fig. 1 | Direct dates for hominins of the Middle to Upper Palaeolithic transition in Eurasia. Directly dated Châtelperronian Neanderthals (blue) and *H. sapiens* (red or black) of the Middle to Upper Palaeolithic transition in Eurasia. The dates from Bacho Kiro Cave (red) are reported in an associated study³, as part of an extensive site chronology. Asterisks mark the dates that were combined using the R_Combine function in OxCal v.4.3. Accelerated mass spectrometry (AMS) laboratory codes are shown in parentheses; all the dates shown here are in Supplementary Table 16, with sample information and references.

and one of the ZooMS-identified hominin bone fragments (AA7-738) are identical, which indicates that these specimens belonged either to the same individual or to two maternally related individuals. In a tree relating these mtDNA genomes to the known mtDNA sequences of 54 present-day humans, 12 ancient *H. sapiens*, 22 Neanderthals, 4 Denisovans and a hominin from Sima de los Huesos, all of the Bacho Kiro Cave mtDNA genomes fall within the variation of *H. sapiens* (Fig. 2, Extended Data Fig. 8). The specimens from layer I yielded mtDNA sequences that fall close to the base of each of the three major macro-haplogroups of present-day non-Africans (M, N and R). Although the mtDNA sequences belong to different macro-haplogroups, they differ (at most) at 15 positions from each other—which is lower than the differences observed among 97.5% of contemporary European individuals who are not closely related to one another²⁶. The older Bacho Kiro population contains early representatives of the macro-haplogroup M, which is not present in Europe today²⁷. Furthermore, the mtDNA genomes of the Bacho Kiro Cave specimens accumulated fewer substitutions than those of present-day humans. Using 10 directly dated ancient *H. sapiens* as calibration points^{28,29} (Supplementary Information section 5), we obtained genetic dates that range from 44,830 to 42,616 yr BP for the layer-I hominins (Extended Data Table 2), in good agreement with the calibrated radiocarbon dates (Fig. 1).

The fauna associated with these *H. sapiens* specimens (11,259 piece plotted animal bone fragments from layers I and J) includes 23 species, dominated by *Bos* or *Bison*, cervids and caprines, alongside equids (Supplementary Information section 6). The species composition comprises a mix of taxa adapted both to cold and to warmer environments, characteristic of the faunal record during marine isotope stage 3 in the Balkans^{30,31}. A variety of carnivores are also present, dominated by cave bear (*Ursus spelaeus*). Zooarchaeological analyses strongly indicate that the accumulation of the fauna is predominantly anthropogenic. One notable aspect of the faunal assemblage is the presence of numerous anthropogenically modified objects (Fig. 3, Supplementary Information section 6): worked pieces include awls, *lissoirs* ('smoothers') and incised pieces. Several of the artefacts have red staining that is consistent with the use of ochre. We identified 1 perforated ivory bead and 12 perforated or grooved pendants, 11 of which were made from cave bear teeth and 1 from an ungulate tooth (Fig. 3).

The stone tools associated with *H. sapiens* in layer I were initially assigned to the Bachokirian techno-complex because they did not fit

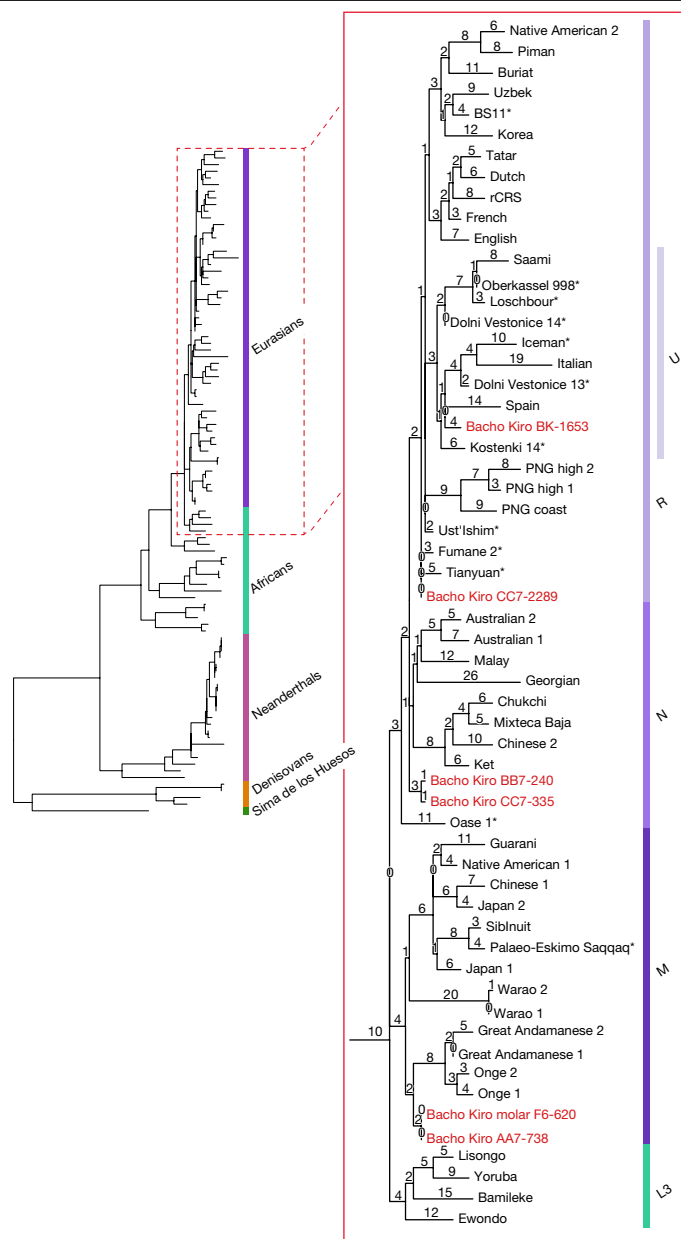


Fig. 2 | Maximum parsimony tree. Maximum parsimony tree relating Bacho Kiro Cave mtDNAs to 54 present-day humans, 12 ancient *H. sapiens*, 22 Neanderthals, 4 Denisovans and 1 individual from Sima de los Huesos. The insert shows the part of the tree closest to the mtDNAs of the specimens from Bacho Kiro Cave. Bacho Kiro Cave mtDNAs are red. Asterisks denote mtDNA from ancient *H. sapiens* (Supplementary Table 9) other than the Bacho Kiro Cave specimens. The number of inferred substitutions per sequence is given above each branch. A chimpanzee mtDNA sequence was used to root the tree (not shown). rCRS, revised Cambridge Reference Sequence. U, R, N, M and L3 refer to the mitochondrial haplogroups.

comfortably with either the Middle Palaeolithic or Aurignacian-like Upper Palaeolithic techno-complexes. We now know that these tools fit within the IUP¹⁵. IUP assemblages—similar to that of Bacho Kiro Cave (Supplementary Information section 2)—are characterized by blades and tool types typical of the Upper Palaeolithic, but with some Levallois forms and faceted platforms that are reminiscent of the preceding Middle Palaeolithic and African Middle Stone Age² (Extended Data Figs. 3, 4). IUP assemblages, which span Eurasia from central Europe to Mongolia, occur before the appearance of Upper Palaeolithic assemblages characterized by bladelet production, and arguably have their origin



Fig. 3 | Bone tools and personal ornaments from Bacho Kiro Cave layers I and J (Niche 1 and Main sectors). a–j, Pendants made from perforated and grooved teeth (a, ungulate; b–j, cave bear). k, l, o, Awls. m, Anthropogenically modified piece. n, p, Lissoirs. q, Ivory bead. Further details are provided in Supplementary Table 15. Scale bars, 1 cm (a–o, q), 3 cm (p).

in southwest Asia (Extended Data Fig. 2). For instance, the Bacho Kiro Cave IUP is similar to the IUP from layers I–F at Üçağızlı Cave (Turkey) in terms of lithic technology, typology, and the presence of shaped bone tools and pendants, as well as with respect to ages^{32,33}.

The Bacho Kiro Cave site clearly demonstrates that the IUP in this region was made by *H. sapiens*, and is consistent with models that attribute the spread of the IUP to the dispersal of our species throughout large parts of Eurasia. The presence of IUP assemblages documents a wave of peopling that precedes the spread of the first Upper Palaeolithic bladelet techno-complexes—such as the Early Ahmarian industry in the Levant, the Early Kozarnikan industry in the eastern Balkans and the Protoaurignacian industry in western and central Europe—by several millennia^{1,34}. At Bacho Kiro Cave, the IUP starts before 45,000 cal. BP and, as the assemblage of the upper part of layer J is identical to that from layer I, it may begin as early as 47,000 cal. BP³. We now have evidence for *H. sapiens* in Eurasia spanning from Ust'-Ishim²⁸ in western Siberia to Bacho Kiro Cave in eastern Europe, directly dated to approximately 45,000 cal. BP. Together, the behavioural and biological evidence strongly suggest a relatively rapid dispersal of IUP assemblages from southwest Asia³⁵ into mid-latitude Eurasia by groups that—contrary to Aurignacian populations—seem unrelated to present-day European populations²⁸. Direct contact with Neanderthals must have occurred much earlier in eastern Europe than in western Europe, where the latest Neanderthals and their associated assemblages persisted until at least about 40,000 cal. BP^{1,5,6}. In Romania, the Peștera cu Oase *H. sapiens*

individual had a Neanderthal ancestor as recently as four-to-six generations back in his family tree³⁶. In light of the Bacho Kiro Cave results, the 42,000–37,000 cal. yr BP radiocarbon age of the Peștera cu Oase fossil implies an extended period of contact between Neanderthals and *H. sapiens* in eastern Europe. Alternatively, it may be that the direct date of Peștera cu Oase—which was obtained before recent improvements in pretreatment techniques—is an underestimate, and that local coexistence was more ancient and ephemeral. The IUP pendants of Bacho Kiro Cave (Fig. 3) are notably similar to artefacts produced by late Neanderthals of the Châtelperronian layers at Grotte du Renne (France)⁴. Whatever the cognitive complexity of the last Neanderthals might have been, the earlier age of the Bacho Kiro Cave material supports the notion that these specific behavioural novelties seen in declining Neanderthal populations resulted from contacts with migrant *H. sapiens*⁷.

Online content

Any methods, additional references, Nature Research reporting summaries, source data, extended data, supplementary information, acknowledgements, peer review information; details of author contributions and competing interests; and statements of data and code availability are available at <https://doi.org/10.1038/s41586-020-2259-z>.

- Hublin, J.-J. The modern human colonization of western Eurasia: when and where? *Quat. Sci. Rev.* **118**, 194–210 (2015).
- Kuhn, S. L. & Zwyns, N. Rethinking the initial Upper Paleolithic. *Quat. Int.* **347**, 29–38 (2014).
- Fewlass, H. et al. A ¹⁴C chronology for the Middle to Upper Palaeolithic transition at Bacho Kiro cave, Bulgaria. *Nat. Ecol. Evol.* <https://doi.org/10.1038/s41559-020-1136-3> (2020).
- White, R. Personal ornaments from the Grotte du Renne at Arcy-sur-Cure. *Athena Review* **2**, 41–46 (2001).
- Welker, F. et al. Palaeoproteomic evidence identifies archaic hominins associated with the Châtelperronian at the Grotte du Renne. *Proc. Natl Acad. Sci. USA* **113**, 11162–11167 (2016).
- Hublin, J.-J. et al. Radiocarbon dates from the Grotte du Renne and Saint-Césaire support a Neanderthal origin for the Châtelperronian. *Proc. Natl Acad. Sci. USA* **109**, 18743–18748 (2012).
- Hublin, J.-J., Spoor, F., Braun, M., Zonneveld, F. & Condemi, S. A late Neanderthal associated with Upper Palaeolithic artefacts. *Nature* **381**, 224–226 (1996).
- Ruebens, K., McPherron, S. J. P. & Hublin, J.-J. On the local Mousterian origin of the Châtelperronian: integrating typo-technological, chronostratigraphic and contextual data. *J. Hum. Evol.* **86**, 55–91 (2015).
- Higham, T. et al. The earliest evidence for anatomically modern humans in northwestern Europe. *Nature* **479**, 521–524 (2011).
- Benazzi, S. et al. Early dispersal of modern humans in Europe and implications for Neanderthal behaviour. *Nature* **479**, 525–528 (2011).
- White, M. & Pettitt, P. Ancient digs and modern myths: the age and context of the Kent's Cavern 4 maxilla and the earliest *Homo sapiens* specimens in Europe. *Eur. J. Archaeol.* **15**, 392–420 (2012).
- Zilhão, J., Banks, W. E., d'Errico, F. & Gioia, P. Analysis of site formation and assemblage integrity does not support attribution of the Uluzzian to modern humans at Grotta del Cavallo. *PLoS ONE* **10**, e0131181 (2015).
- Kozłowski, J. K. *Excavation in the Bacho Kiro Cave (Bulgaria): Final Report*. 172 (Panstwowe Wydawnictwo Naukowe, 1982).
- Hedges, R. E. M., Housley, R. A., Bronk Ramsey, C. & Klinken, G. J. V. Radiocarbon dates from the Oxford AMS system: archaeometry datelist 18. *Archaeometry* **36**, 337–374 (1994).
- Tsanova, T. & Bordes, J. G. in *The Humanized Mineral World: Towards Social and Symbolic Evaluation of Prehistoric Technologies in South Eastern Europe (Proceedings of the ESF Workshop)* (eds Tsonev, T. S. & Montagnari Kokclj, E.) 41–50 (ERAUL, 2003).
- Bailey, S. E. A closer look at Neanderthal postcanine dental morphology: the mandibular dentition. *Anat. Rec.* **269**, 148–156 (2002).
- Bailey, S. E., Skinner, M. M. & Hublin, J.-J. What lies beneath? An evaluation of lower molar trigonid crest patterns based on both dentine and enamel expression. *Am. J. Phys. Anthropol.* **145**, 505–518 (2011).
- Keith, A. Problems relating to the teeth of the earlier forms of prehistoric man. *Proc. R. Soc. Med.* **6**, 103–124 (1913).
- Shaw, J. *The Teeth, the Bony Palate and the Mandible in Bantu Races of South Africa* (Bale and Danielsson, London, 1938).
- Kallay, J. in *Dental Anthropology* (ed. Brothwell, D.) 75–86 (Pergamon, 1963).
- Buckley, M., Collins, M., Thomas-Oates, J. & Wilson, J. C. Species identification by analysis of bone collagen using matrix-assisted laser desorption/ionisation time-of-flight mass spectrometry. *Rapid Commun. Mass Spectrom.* **23**, 3843–3854 (2009).
- Dabney, J. et al. Complete mitochondrial genome sequence of a Middle Pleistocene cave bear reconstructed from ultrashort DNA fragments. *Proc. Natl Acad. Sci. USA* **110**, 15758–15763 (2013).
- Korlević, P. et al. Reducing microbial and human contamination in DNA extractions from ancient bones and teeth. *Biotechniques* **59**, 87–93 (2015).
- Gansauge, M.-T. et al. Single-stranded DNA library preparation from highly degraded DNA using T4 DNA ligase. *Nucleic Acids Res.* **45**, e79 (2017).
- Fu, Q. et al. DNA analysis of an early modern human from Tianyuan Cave, China. *Proc. Natl Acad. Sci. USA* **110**, 2223–2227 (2013).
- Lippold, S. et al. Human paternal and maternal demographic histories: insights from high-resolution Y chromosome and mtDNA sequences. *Investig. Genet.* **5**, 13 (2014).
- Kivisild, T. Maternal ancestry and population history from whole mitochondrial genomes. *Investig. Genet.* **6**, 3 (2015).
- Fu, Q. et al. Genome sequence of a 45,000-year-old modern human from western Siberia. *Nature* **514**, 445–449 (2014).
- Fu, Q. et al. A revised timescale for human evolution based on ancient mitochondrial genomes. *Curr. Biol.* **23**, 553–559 (2013).
- van der Made, J. in *The Encyclopedia of Archaeological Sciences* (ed. López Varela, S. L.) 1–4 (Wiley-Blackwell, 2018).
- Guérin, C. Première biozonation du Pléistocène Européen, principal résultat biostratigraphique de l'étude des Rhinocerotidae (Mammalia, Perissodactyla) du Miocène terminal au Pléistocène supérieur d'Europe Occidentale. *Geobios* **15**, 593–598 (1982).
- Kuhn, S. L. et al. The early Upper Paleolithic occupations at Uçağızlı Cave (Hatay, Turkey). *J. Hum. Evol.* **56**, 87–113 (2009).
- Kuhn, S. L. Upper Paleolithic raw material economies at Uçağızlı cave, Turkey. *J. Anthropol. Archaeol.* **23**, 431–448 (2004).
- Müller, U. C. et al. The role of climate in the spread of modern humans into Europe. *Quat. Sci. Rev.* **30**, 273–279 (2011).
- Hershkovitz, I. et al. Levantine cranium from Manot Cave (Israel) foreshadows the first European modern humans. *Nature* **520**, 216–219 (2015).
- Fu, Q. et al. An early modern human from Romania with a recent Neanderthal ancestor. *Nature* **524**, 216–219 (2015).

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

© The Author(s), under exclusive licence to Springer Nature Limited 2020

Methods

No statistical methods were used to predetermine sample size.

Excavation methods

The site was excavated closely following existing protocols^{37–42}. Layers were defined first on lithological criteria, and second on archaeological criteria. Our stratigraphy was determined and named independently of previous excavations¹³. Additionally, we excavated in two unconnected areas of the site, and thus separate naming conventions were used between these two areas. The excavations in the south area (Extended Data Fig. 1) are known as the Main sector, and the layers are named with letters for the large divisions and numbers for divisions within these (for example, layer I or layer AO). The other area excavated is a niche to the east of the previous excavations. We call this area Niche 1 (or N1), and all layer names from this area are prefixed with N1-. Where we hypothesize a link between the two sectors, we use the same layer name (for example, layers N1-I and I). Where we are unable to form a strong hypothesis about the link between the two sectors, we use different layer names (numbers in this case) in the Niche 1 sector to denote this (for example, N1-3), followed by letter for internal subdivisions (for example, N1-3a). All finds were recorded by layer and 3D coordinates (using an arbitrary grid established for the excavation and aligned to the previous excavations) measured with Leica total stations (5'' accuracy) using data collectors with self-authored software (EDM-Mobile). All lithics and fauna >20 mm in length and all specialists' samples (for example, ancient DNA, micromorphology, phytoliths and so on) were provenienced and given unique identifiers (IDs). Complete bones, identifiable teeth and human remains <20 mm in length (but larger than microfauna) were also given coordinates and IDs. Natural stones >10 cm in length were recorded with a single coordinate, and stones >20 cm in length were measured with multiple coordinates to describe their volume and orientation. The sediment, excluding recorded stones and artefacts, was collected by 9-l buckets and wet-screened on-site through 6- and 1.2-mm meshes to form two fractions. Buckets have unique IDs. Their coordinates were measured first in the centre of the area to be excavated and then again at the centre of the area excavated at the completion of the bucket. Large (>20-mm-long) objects found in the sediment in the buckets during wet-screening were given IDs and assigned the coordinates of the bucket. All features were provenienced. Digital photographs documenting the excavation were recorded daily, and final sections were documented through a combination of digital photography, drawing, and total station measures. Additionally, structure-from-motion models were made of all final sections and excavation areas. These models were georeferenced to the excavation grid using total station coordinates.

ZooMS

We screened 1,271 fragmentary bone and tooth specimens from Bacho Kiro Cave using ZooMS²¹. Eleven bone specimens were derived from the previous excavations at the site¹³, 371 from our excavations in the Main sector, and 889 bone specimens from the Niche 1 area. We particularly focused on IUP layers I and N1-I ($n = 822$). Extraction and analytical protocols followed previously published work⁵. In brief, a small bone sample (<20 mg) was taken from each bone or dentine specimen. The sample was incubated at 65 °C for an hour in 50 mM ammonium-bicarbonate buffer, digested overnight using trypsin (Promega) at 37 °C, acidified using 20% TFA, and cleaned on C18 ZipTips (either from Sigma-Aldrich or Thermo Scientific). MALDI-TOF MS analysis was conducted at the IZI Fraunhofer in Leipzig⁴³. MALDI-TOF MS spectra were analysed in comparison to a reference database containing collagen-peptide marker masses of all medium- to larger-sized genera in existence in western Eurasia during the Late Pleistocene epoch⁵. In cases in which ammonium-bicarbonate extraction failed, an attempt was made to recover further informative collagen peptides through acid

demineralization of the same bone sample, as previously explained⁵. Collagen deamidation in these spectra was assessed for two peptides (P1105 and P1706)^{44,45}.

Bone pretreatment and accelerator mass spectrometry dating

Small aliquots (80–110 mg) of the six ZooMS-identified hominin bone fragments were sampled for dating to preserve as much material as possible for further analyses. Collagen was extracted using a previously described technique⁴⁶ for small bone sample sizes, based on a modified Longin collagen-extraction protocol⁴⁷ followed by an ultrafiltration step⁴⁸. In brief, the outer surfaces of the bone samples were removed with a sandblaster, and samples were removed using a rotary tool. The bones were demineralized in 0.5 M HCl at 4 °C until soft and CO₂ effervescence had stopped. Then, 0.1 M NaOH was added for 10 min at room temperature to remove humic acid contamination, and samples were re-acidified in 0.5 M HCl. The collagen was gelatinized in acidic water (HCl pH 3) at 70 °C for several hours (4–6 h). The collagen samples were then passed through an Eze Filter (Elkay Laboratories) to remove large particles (>80 µm) and separated by molecular weight with pre-cleaned Sartorius VivaSpin Turbo 15 ultrafilters (30 kDa molecular weight cut-off (MWCO))^{49,50}. The samples were freeze-dried and the large molecular fraction (>30 kDa) was graphitized using Automated Graphitisation Equipment III⁵¹ and measured using the latest model of the MICADAS accelerator mass spectrometer⁵² in the Laboratory of Ion Beam Physics at ETH-Zurich (laboratory code ETH). Small aliquots (66–89 mg) of a background cave bear bone (>50,000 yr BP) were extracted alongside the samples to monitor contamination introduced in the laboratory⁵³. These were measured in the same magazine as the hominin samples and used in the age calculation. Oxalic acid II standards were also measured in the same magazine and used for normalization. Data reduction was performed using BATS software⁵⁴. An additional 1‰ was added to the error calculation of the samples, as per standard practice. The dates were calibrated using the IntCal13⁵⁵ dataset in OxCal v.4.3⁵⁶.

Shape analysis of the molar enamel–dentine junction

Enamel and dentine tissues (Extended Data Fig. 5) of lower second molars were segmented using the 3D voxel value histogram and its distribution of greyscale values^{57,58}. After segmentation, the enamel–dentine junction was reconstructed as a triangle-based surface model using Avizo. Small enamel–dentine junction defects were corrected digitally using the 'fill holes' module of Geomagic Studio. We then used Avizo to digitize 3D landmarks and curve-semilandmarks on the enamel–dentine junction surface^{57,58}. Anatomical landmarks were placed on the tip of the dentine horn of the protoconid, metaconid, entoconid and hypoconid. A sequence of landmarks was also placed along the marginal ridge connecting the dentine horns, beginning at the top of the protoconid and moving in lingual direction; the points along this ridge curve were then later resampled to the same point count on every specimen using Mathematica. Likewise, we digitized and resampled a curve along the cemento–enamel junction as a closed curve starting and ending below the protoconid horn and the mesiobuccal corner of the cervix. The resampled points along the two ridge curves were subsequently treated as sliding curve semilandmarks and analysed using geometric morphometrics together with the four anatomical landmarks. Landmarks not preserved on the Bacho Kiro Cave specimen were removed before principal component analysis. The specimens of *Homo erectus* sensu lato include KNM-ER 1802, KNM-ER 992 and Sangiran 1b. Specimens of archaic Middle Pleistocene hominins include Balanica 1, Mauer, Xiahe and KNM-ER BK 67. The Neanderthal sample includes Abri Suard S36, Krapina 1, 6, 9, 53, 54, 55, 57, 59, 80, 86, 105 and 107, La Quina H9, Le Moustier 1, Regourdou, Scladina 4A1, El Sidron 540 and 755, and Vindija 11.39. The fossil *H. sapiens* sample includes Dar es Soltane II H4, El Harhoura, Jebel Irhoud 3 and 11, Qafzeh 9, 10, 11 and 15, and Temara. The recent *H. sapiens* sample includes clinical extractions

Article

from dentists based in Germany, Neolithic specimens from Belgium (Royal Belgian Institute of Natural Sciences) and specimens from the Francisc J. Rainer Collection (Institutul de Antropologie ‘Francisc J. Rainer’).

DNA extraction and library preparation

Samples of between 29.3 mg and 64.7 mg of tooth or bone powder were removed from 7 Bacho Kiro Cave specimens (F6-620, AA7-738, BB7-240, CC7-2289, CC7-335, F6-597 and BK-1653) using a sterile dentistry drill (Supplementary Table 4) after a thin layer of surface was removed from the sampling areas. DNA was extracted from the powder using a silica-based method²² as previously described²³. Five single-stranded DNA libraries²⁴ were made from 10 µl from each extract on an automated liquid handling platform (Bravo NGS workstation B, Agilent Technologies)⁵⁹. A control oligonucleotide was spiked into each reaction to determine the efficiency of library preparation⁶⁰, and quantitative PCR was used to determine the total number of unique library molecules as well as the number of oligonucleotides that were successfully converted^{24,60}. The libraries were amplified into plateau with AccuPrime Pfx DNA polymerase (Life Technologies)⁶¹ and labelled with two unique indices^{23,62}. Half of the volume of the amplified libraries (50 µl) was purified using SPRI beads on the automated liquid handling platform⁵⁹. The concentrations of the purified DNA libraries were determined using a NanoDrop Spectrophotometer (NanoDrop Technologies).

mtDNA capture and sequencing

An aliquot of each amplified library was enriched for human mtDNA using a bead-based hybridization method²⁹. Enriched libraries were sequenced on an Illumina MiSeq platform in a double index configuration (2 × 76 cycles)⁶² and base-calling was done using Bustard (Illumina). Overlapping paired-end reads were merged into single sequences and the adapters were trimmed using leeHom⁶³. The Burrows–Wheeler Aligner (BWA, version: 0.5.10-*evan.9-1-g44db244*; <https://github.com/mpieva/network-aware-bwa>)⁶⁴, with parameters adjusted for ancient DNA (‘-n 0.01 -o 2 -l 16500’)⁶⁵, was used to align the data to the revised Cambridge Reference Sequence (NC_01290). Only reads with perfect matches to the expected index combinations were retained for downstream analyses. PCR duplicates were removed using bam-rmdup (version 0.6.3; <https://bitbucket.org/ustenzel/biohazard>). SAMtools (version 1.3.1)⁶⁶ was used to filter for fragments that were longer than 35 base pairs and that had a mapping quality of at least 25. We merged the libraries originating from the same extract (that is, the same specimen) using SAMtools merge to produce the final dataset.

Phylogenetic inferences

We reconstructed the mitochondrial genomes of the Bacho Kiro Cave specimens once by using all mapped fragments longer than 35 base pairs with a mapping quality of at least 25 and once using only fragments with a cytosine (C) to thymine (T) difference to the reference genome at the first three and/or last three terminal positions³⁶ (that is, putatively deaminated fragments). We called a consensus base at each position along the mtDNA that was covered by at least 3 DNA fragments and at which at least 2/3 of fragments carried an identical base and the base quality was 20 or higher⁶⁷. To prevent deamination-induced substitutions affecting the calling of a consensus base, we converted A on the reverse strands and T on the forward strands in the first three and the last three positions of a fragment into N.

The libraries prepared from the F6-597 specimen yielded too few informative mtDNA fragments to reconstruct a complete mtDNA using putatively deaminated fragments. We investigated the state of F6-597 DNA fragments that overlapped positions ‘diagnostic’ for each branch in a mtDNA tree relating present-day humans, Neanderthals, Denisovans and the hominin from Sima de los Huesos⁶⁸ (Supplementary Table 6). To diminish the influence of substitutions derived from deamination, all forward strands were ignored if one of the possible states

at an informative state was a C and all reverse strands were ignored if one of the possible states was G.

We aligned the reconstructed mitochondrial genomes of the Bacho Kiro Cave individuals to the mtDNA genomes of 54 present-day humans from a wide geographical distribution⁶⁹, 12 ancient *H. sapiens*^{25,28,41,70–73} (Supplementary Table 9), 22 Neanderthals^{69,74–78}, 4 Denisovans^{79–82}, a Sima de los Huesos individual⁶⁷ and a chimpanzee⁸³ using MAFFT v.7.271⁸⁴. The number of pairwise differences among the genomes was calculated using MEGA7⁸⁵ and a maximum parsimony tree was reconstructed using Parsimony ratchet as implemented in the R package phangorn⁸⁶. We identified the haplogroup of each of the reconstructed mitochondrial genomes with HaploGrep⁸⁷ based on the PhyloTree database (PhyloTree.org, build 17).

Contamination estimates

We used two complementary approaches to estimate levels of present-day human mtDNA contamination in the libraries. We identified positions at which each of the reconstructed Bacho Kiro Cave mtDNAs differ from at least 99% of a world-wide panel of 311 present-day human mtDNAs^{36,69} (Supplementary Tables 7, 8). We then counted DNA fragments that overlap these positions and did not match the consensus base of the respective specimen, again taking into account the strand orientation in cases in which one of the possible states at an informative site was C or G. In the second approach, we used an iterative probabilistic method, *schmutzi*⁸⁸, which uses a nonredundant database of human mitochondrial genomes to estimate levels of present-day human DNA contamination (Supplementary Information section 5) (parameters: ‘--notusepredC --uselength’).

Molecular DNA dating

We estimated the tip dates of the reconstructed Bacho Kiro Cave mtDNAs using the Bayesian phylogenetic method as implemented in BEAST2 (version 2.4.8)⁸⁹ by aligning the reconstructed mitochondrial genomes to 54 present-day humans and 10 directly radiocarbon-dated ancient *H. sapiens*^{28,29,72,73,80}, which were used for tip calibration. The Neanderthal mtDNA genome of Vindija 33.16⁶⁹ was used as an outgroup. The best-fitting substitution model was determined using jModelTest2⁹⁰. We investigated a strict clock and an uncorrelated log-normal relaxed clock as two models of rate variation and a constant population size and a Bayesian skyline as tree priors²⁸. For each model, we carried out Markov chain Monte Carlo runs with 30,000,000 iterations and sampling every 1,000 steps. After discarding 10% of the iterations as burn-in, the output was analysed with Tracer v.1.5.0 (<http://tree.bio.ed.ac.uk/software/tracer/>). A marginal likelihood estimation⁹¹ analysis was used for model comparison and best support assessment. Both the maximum parsimony and the BEAST2 tree were visualized with FigTree (version v.1.4.2) (<http://tree.bio.ed.ac.uk/software/figtree/>).

Micromorphology

Field observations of the sediments were complemented by archaeological micromorphology analyses. Micromorphological samples were collected as undisturbed blocks by carefully carving and wrapping them with either pre-plastered bandages or soft paper and tape. Thin sections were manufactured by Spectrum Petrographics through a standard procedure of drying the blocks in an oven for several days at about 60 °C. The blocks were then impregnated with a mixture of polyester resin and styrene, to which a catalyst was added. Thin sections were ground to a thickness of 30 µm and observed under a petrographic microscope in plane- and cross-polarized light at magnifications ranging from 20× to 400×. Micromorphological nomenclature follows previously published work^{92,93}.

Reporting summary

Further information on research design is available in the Nature Research Reporting Summary linked to this paper.

Data availability

The data that support the findings of this study are available from the corresponding author upon reasonable request. Genetic sequence reads from all libraries and corresponding negative controls are deposited at European Nucleotide Archive under the study accession number PRJEB35466. The FASTA files of the mitochondrial genomes are deposited in GenBank with the accession numbers MN706602–MN706607. Details are as follows: Bacho Kiro AA7-738, MN706602; Bacho Kiro BB7-240, MN706603; Bacho Kiro BK-1653, MN706604; Bacho Kiro CC7-335, MN706605; Bacho Kiro CC7-2289, MN706606; and Bacho Kiro molar F6-620, MN706607.

37. Dibble, H. L. & Lenoir, M. *The Middle Paleolithic Site of Combe-Capelle Bas (France)* (The University Museum Press, 1995).
38. Turq, A. et al. in *Les Sociétés du Paléolithique dans un Grand Sud-ouest de la France: Nouveaux Gisements, Nouveaux Résultats, Nouvelles Méthodes* (eds. Jaubert, J. et al.) 83–94 (Mémoire de la Société Préhistorique Française, 2008).
39. Chase, P. G., Debénath, A., Dibble, H. L. & McPherron, S. P. in *The Cave of Fontéchevade: Recent Excavations and their Paleoanthropological Implications* (eds Chase, P. G. et al.) 28–62 (Cambridge Univ. Press, 2009).
40. Soressi, M. et al. Neandertals made the first specialized bone tools in Europe. *Proc. Natl Acad. Sci. USA* **110**, 14186–14190 (2013).
41. Richter, D. et al. The age of the hominin fossils from Jebel Irhoud, Morocco, and the origins of the Middle Stone Age. *Nature* **546**, 293–296 (2017).
42. Sandgathe, D. M., Dibble, H. L., McPherron, S. J. P. & Goldberg, P. in *The Middle Paleolithic Site of Pech de l’Azé IV Cave and Karst Systems of the World* (eds Dibble, H. L. et al.) 1–19 (Springer, 2018).
43. Sinet-Mathiot, V. et al. Combining ZooMS and zooarchaeology to study Late Pleistocene hominin behaviour at Fumane (Italy). *Sci. Rep.* **9**, 12350 (2019).
44. Wilson, J., van Doorn, N. L. & Collins, M. J. Assessing the extent of bone degradation using glutamine deamidation in collagen. *Anal. Chem.* **84**, 9041–9048 (2012).
45. Welker, F. et al. Variations in glutamine deamidation for a Châtelperronian bone assemblage as measured by peptide mass fingerprinting of collagen. *Sci. Technol. Archaeol. Res.* **3**, 15–27 (2017).
46. Fewlass, H. et al. Pretreatment and gaseous radiocarbon dating of 40–100 mg archaeological bone. *Sci. Rep.* **9**, 5342 (2019).
47. Longin, R. New method of collagen extraction for radiocarbon dating. *Nature* **230**, 241–242 (1971).
48. Brown, T. A., Nelson, D. E., Vogel, J. S. & Southon, J. R. Improved collagen extraction by modified Longin method. *Radiocarbon* **30**, 171–177 (1988).
49. Bronk Ramsey, C., Higham, T., Bowles, A. & Hedges, R. Improvements to the pretreatment of bone at Oxford. *Radiocarbon* **46**, 155–163 (2004).
50. Brock, F., Bronk Ramsey, C. & Higham, T. Quality assurance of ultrafiltered bone dating. *Radiocarbon* **49**, 187–192 (2007).
51. Wacker, L., Némec, M. & Bourquin, J. A revolutionary graphitisation system: fully automated, compact and simple. *Nucl. Instrum. Methods Phys. Res. B* **268**, 931–934 (2010).
52. Wacker, L. et al. MICADAS: routine and high-precision radiocarbon dating. *Radiocarbon* **52**, 252–262 (2010).
53. Korlević, P., Talamo, S. & Meyer, M. A combined method for DNA analysis and radiocarbon dating from a single sample. *Sci. Rep.* **8**, 4127 (2018).
54. Wacker, L., Christl, M. & Sina, H. A. Bats: a new tool for AMS data reduction. *Nucl. Instrum. Methods Phys. Res. B* **268**, 976–979 (2010).
55. Reimer, P. J. et al. IntCal13 and Marine13 radiocarbon age calibration curves 0–50,000 years cal BP. *Radiocarbon* **55**, 1869–1887 (2013).
56. Bronk Ramsey, C. Bayesian analysis of radiocarbon dates. *Radiocarbon* **51**, 337–360 (2009).
57. Skinner, M. M., Gunz, P., Wood, B. A. & Hublin, J.-J. Enamel–dentine junction (EDJ) morphology distinguishes the lower molars of *Australopithecus africanus* and *Paranthropus robustus*. *J. Hum. Evol.* **55**, 979–988 (2008).
58. Skinner, M. M., Gunz, P., Wood, B. A., Boesch, C. & Hublin, J.-J. Discrimination of extant *Pan* species and subspecies using the enamel–dentine junction morphology of lower molars. *Am. J. Phys. Anthropol.* **140**, 234–243 (2009).
59. Slon, V. et al. Neandertal and Denisovan DNA from Pleistocene sediments. *Science* **356**, 605–608 (2017).
60. Glocke, I. & Meyer, M. Extending the spectrum of DNA sequences retrieved from ancient bones and teeth. *Genome Res.* **27**, 1230–1237 (2017).
61. Dabney, J. & Meyer, M. Length and GC-biases during sequencing library amplification: a comparison of various polymerase–buffer systems with ancient and modern DNA sequencing libraries. *Biotechniques* **52**, 87–94 (2012).
62. Kircher, M., Sawyer, S. & Meyer, M. Double indexing overcomes inaccuracies in multiplex sequencing on the Illumina platform. *Nucleic Acids Res.* **40**, e3 (2012).
63. Renaud, G., Stenzel, U. & Kelso, J. leeHom: adaptor trimming and merging for Illumina sequencing reads. *Nucleic Acids Res.* **42**, e141 (2014).
64. Li, H. & Durbin, R. Fast and accurate long-read alignment with Burrows–Wheeler transform. *Bioinformatics* **26**, 589–595 (2010).
65. Meyer, M. et al. A high-coverage genome sequence from an archaic Denisovan individual. *Science* **338**, 222–226 (2012).
66. Li, H. et al. The Sequence Alignment/Map format and SAMtools. *Bioinformatics* **25**, 2078–2079 (2009).
67. Meyer, M. et al. A mitochondrial genome sequence of a hominin from Sima de los Huesos. *Nature* **505**, 403–406 (2014).
68. Meyer, M. et al. Nuclear DNA sequences from the Middle Pleistocene Sima de los Huesos hominins. *Nature* **531**, 504–507 (2016).
69. Green, R. E. et al. A complete Neandertal mitochondrial genome sequence determined by high-throughput sequencing. *Cell* **134**, 416–426 (2008).
70. Benazzi, S. et al. The makers of the Protoaurignacian and implications for Neandertal extinction. *Science* **348**, 793–796 (2015).
71. Ermini, L. et al. Complete mitochondrial genome sequence of the Tyrolean Iceman. *Curr. Biol.* **18**, 1687–1693 (2008).
72. Gilbert, M. T. P. et al. Paleo-Eskimo mtDNA genome reveals matrilineal discontinuity in Greenland. *Science* **320**, 1787–1789 (2008).
73. Krause, J. et al. A complete mtDNA genome of an early modern human from Kostenki, Russia. *Curr. Biol.* **20**, 231–236 (2010).
74. Green, R. E. et al. A draft sequence of the Neandertal genome. *Science* **328**, 710–722 (2010).
75. Posth, C. et al. Deeply divergent archaic mitochondrial genome provides lower time boundary for African gene flow into Neanderthals. *Nat. Commun.* **8**, 16046 (2017).
76. Prüfer, K. et al. The complete genome sequence of a Neanderthal from the Altai Mountains. *Nature* **505**, 43–49 (2014).
77. Rougier, H. et al. Neandertal cannibalism and Neandertal bones used as tools in Northern Europe. *Sci. Rep.* **6**, 29005 (2016).
78. Skoglund, P. et al. Separating endogenous ancient DNA from modern day contamination in a Siberian Neandertal. *Proc. Natl Acad. Sci. USA* **111**, 2229–2234 (2014).
79. Krause, J. et al. The complete mitochondrial DNA genome of an unknown hominin from southern Siberia. *Nature* **464**, 894–897 (2010).
80. Reich, D. et al. Genetic history of an archaic hominin group from Denisova Cave in Siberia. *Nature* **468**, 1053–1060 (2010).
81. Sawyer, S. et al. Nuclear and mitochondrial DNA sequences from two Denisovan individuals. *Proc. Natl Acad. Sci. USA* **112**, 15696–15700 (2015).
82. Slon, V. et al. A fourth Denisovan individual. *Sci. Adv.* **3**, e1700186 (2017).
83. Horai, S. et al. Man’s place in Hominoida revealed by mitochondrial DNA genealogy. *J. Mol. Evol.* **35**, 32–43 (1992).
84. Katoh, K. & Standley, D. M. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Mol. Biol. Evol.* **30**, 772–780 (2013).
85. Kumar, S., Stecher, G. & Tamura, K. MEGA7: molecular evolutionary genetics analysis version 7.0 for bigger datasets. *Mol. Biol. Evol.* **33**, 1870–1874 (2016).
86. Schliep, K. P. phangorn: phylogenetic analysis in R. *Bioinformatics* **27**, 592–593 (2011).
87. Kloss-Brandstätter, A. et al. HaploGrep: a fast and reliable algorithm for automatic classification of mitochondrial DNA haplogroups. *Hum. Mutat.* **32**, 25–32 (2011).
88. Renaud, G., Slon, V., Duggan, A. T. & Kelso, J. Schmutzi: estimation of contamination and endogenous mitochondrial consensus calling for ancient DNA. *Genome Biol.* **16**, 224 (2015).
89. Bouckaert, R. et al. BEAST 2: a software platform for Bayesian evolutionary analysis. *PLOS Comput. Biol.* **10**, e1003537 (2014).
90. Darriba, D., Taboada, G. L., Doallo, R. & Posada, D. jModelTest 2: more models, new heuristics and parallel computing. *Nat. Methods* **9**, 772 (2012).
91. Baele, G. et al. Improving the accuracy of demographic and molecular clock model comparison while accommodating phylogenetic uncertainty. *Mol. Biol. Evol.* **29**, 2157–2167 (2012).
92. Stoops, G. *Guidelines for Analysis and Description of Soil and Regolith Thin Sections*. 184 (Soil Science Society of America, 2003).
93. Courty, M. A., Goldberg, P. & Macphail, R. *Soils and Micromorphology in Archaeology* 344 (Cambridge Univ. Press, 1989).

Acknowledgements We thank the tourism association of Bacho Kiro Cave in the town of Dryanovo, the History museum – Dryanovo, the Regional History museum in the city of Gabrovo, Dryanovo town hall and V. Lafchiiski for their assistance with the fieldwork and in the laboratory; N. Spassov from the National Museum of Natural History in Sofia for cooperating and hosting researchers of our project; H. Temming and J. Honeyford for their technical assistance and S. Nagel, B. Nickel, B. Schellbach and A. Weihmann for their help with the ancient DNA laboratory procedures and sequencing. Field operations were funded by the Max Planck Society. AixMICADAS and its operation are funded by Collège de France and the EQUIPEX ASTER-CEREGE (principal investigator, E.B.). S.T. is funded by the European Research Council under the European Union’s Horizon 2020 Research and Innovation Programme (grant agreement no. 803147-951 RESOLUTION). The ancient DNA part of this study was funded by the Max Planck Society and the European Research Council (grant agreement no. 694707 to S.P.).

Author contributions J.-J.H. designed the study. T. Tsanova, N.S., V.A., S.S., R.S., E.E., Z.R. and S.P.M. collected field data; H.F., B.K., L.W., E.B., Y.F., T. Tuna and S.T. established the radiocarbon dates; V.A. studied the micromorphology of the sediments; S.B., M.M.S. and J.-J.H. analysed hominin dental morphology; V.S.-M., L.P., F.W. and A.W. performed ZooMS; M.H., M.M. and S.P. performed mtDNA analysis; T. Tsanova, N.S., N.Z., S.S., I.K., V.D., J.M. and S.P.M. conducted the study of the lithics; G.M.S., R.S., V.P. and N.L.M. analysed the faunal assemblages and the osseous objects. J.-J.H. wrote the paper with contributions of all authors.

Competing interests The authors declare no competing interests.

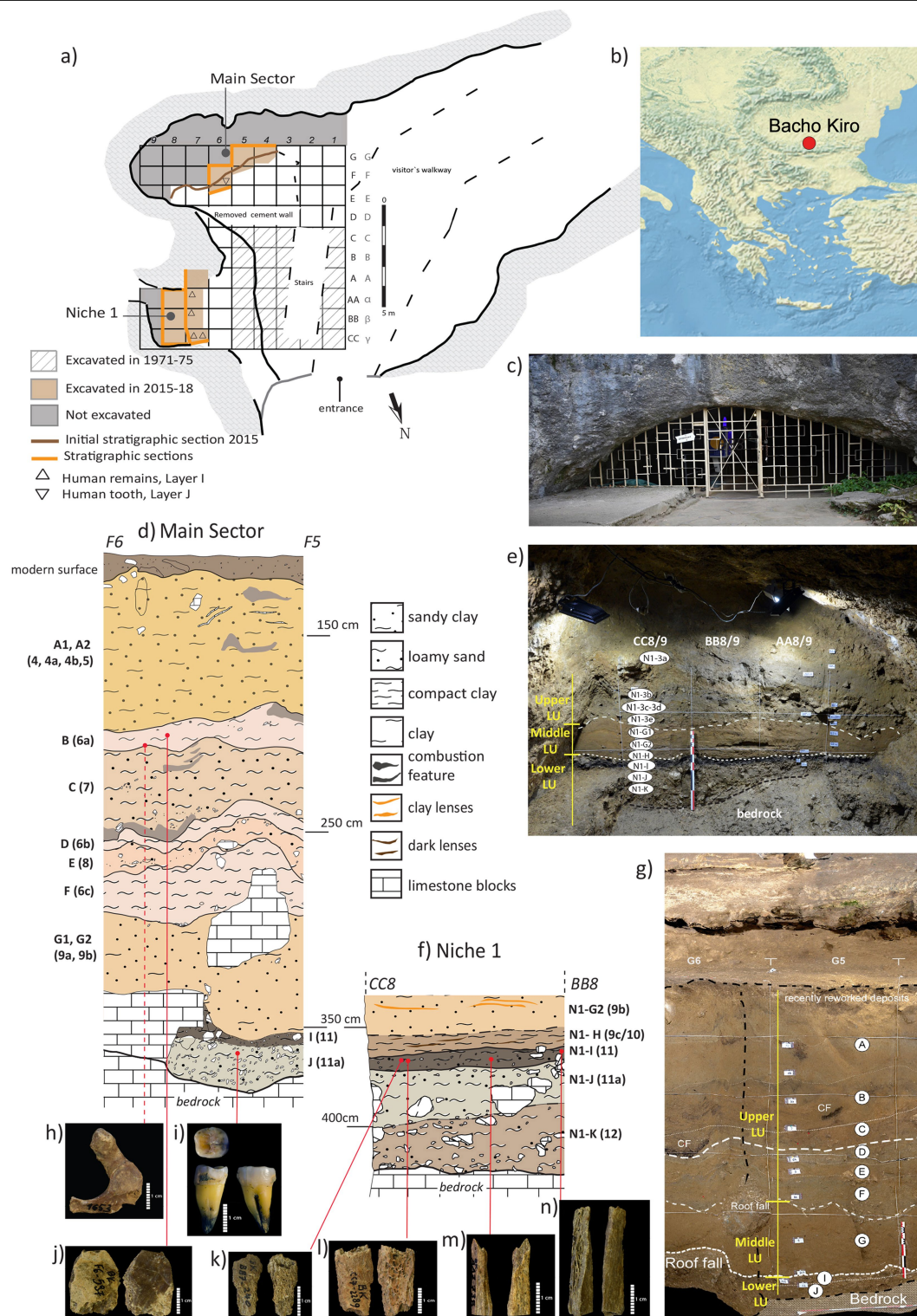
Additional information

Supplementary information is available for this paper at <https://doi.org/10.1038/s41586-020-2259-z>.

Correspondence and requests for materials should be addressed to J.-J.H.

Peer review information *Nature* thanks William Banks, Richard G. Klein and the other, anonymous, reviewer(s) for their contribution to the peer review of this work.

Reprints and permissions information is available at <http://www.nature.com/reprints>.

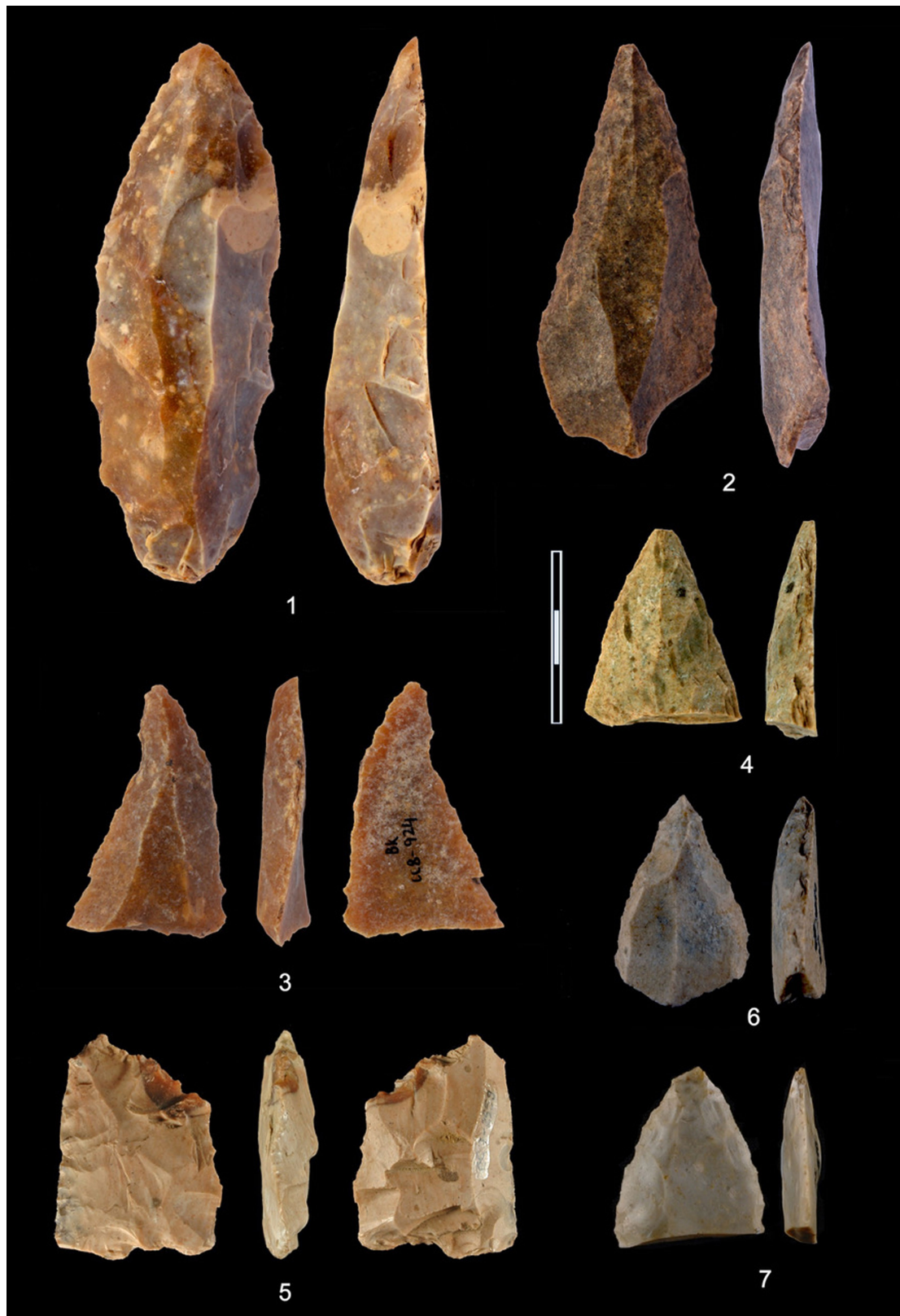


Extended Data Fig. 1 | Excavations at Bacho Kiro Cave, 2015–2018. **a**, Plan view of the entrance and the excavated areas of the cave, with the grid system of our recent excavations (letters in the left column) and those of the 1971–1975 excavations (letters in the right column). **b**, Site location in southeastern Europe. **c**, Photograph of the entrance of the cave. The floor is artificially raised; the original entrance was several metres lower than shown in this photograph. **d**, Initial stratigraphic section drawing of the exposed profile from the Main sector in 2015 (codes for the archaeological layers are on the left, with the corresponding layers from the 1971–1975 excavations in parentheses). **e**, Frontal view of the Niche 1 sector and its stratigraphic subdivisions. **f**, Lower

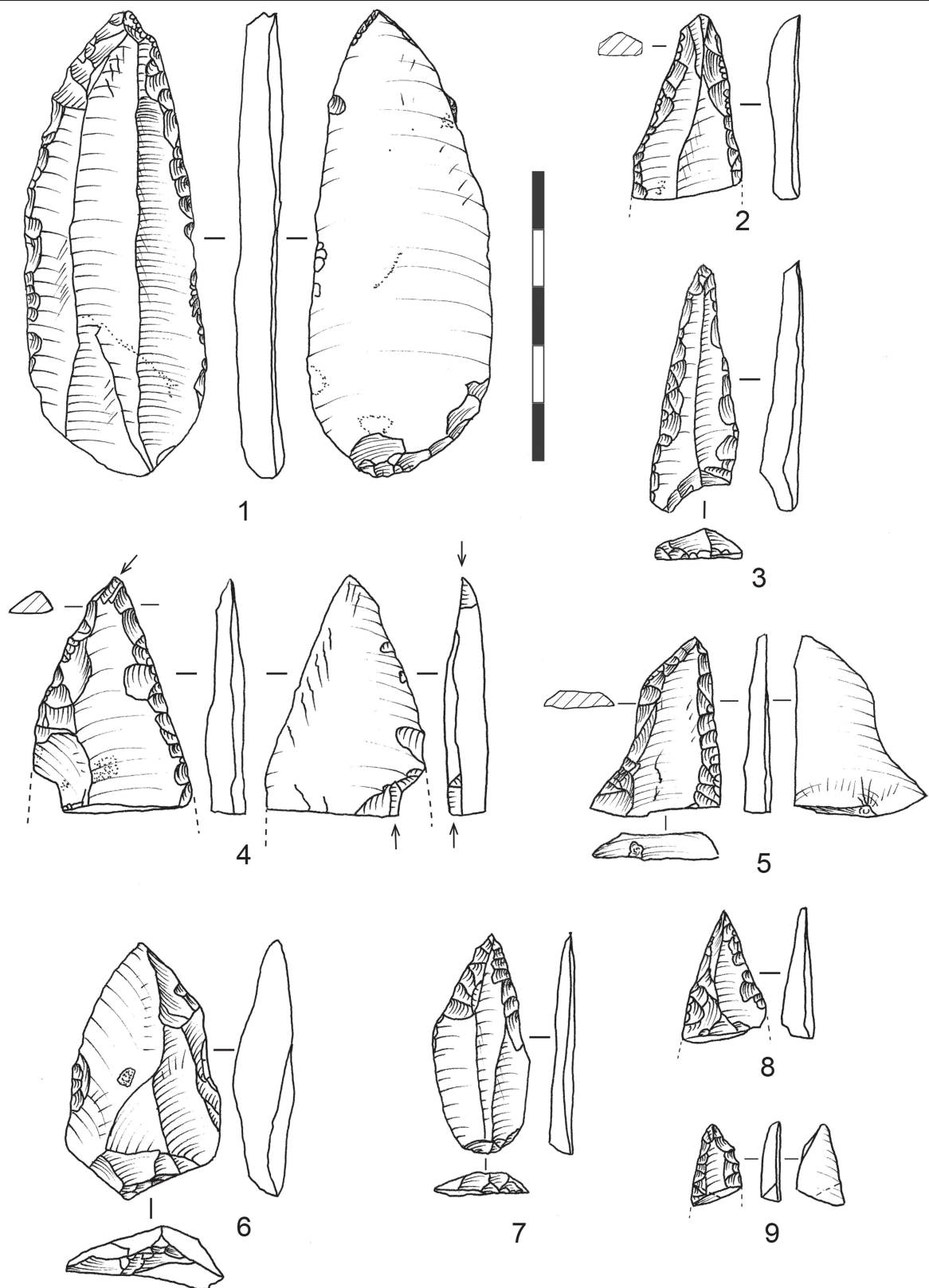
part of the stratigraphic section drawing of the Niche 1 sector, in 2018. Note the thickness and preservation of the lower deposits here in comparison with the Main sector profile. **g**, Photograph of the Main sector transversal section on the line between squares F5–F6 and squares G5–G6 before excavation in 2015. CF, combustion feature. **h–n**, Hominin remains identified by ZooMS with their IDs: BK-1653 (**h**) and F6-597 (**j**) from layer B, with **h** coming from the 1971–1975 excavations (dashed line); BB7-240 (**k**), CC7-2289 (**l**), CC7-335 (**m**) and AA7-738 (**n**) from layer N1-I. Continuous lines connect the fossils with their find locations. **i**, Second lower molar (F6-620) from layer J in the Main sector.



Extended Data Fig. 2 | Geographical distributions. Geographical distribution of the main IUP sites of western and central Eurasia (black dots), directly dated early *H. sapiens* predating 37,000 cal. BP (empty black dots) and directly dated late Neanderthals associated with Châtelperronian assemblages (orange squares). Bacho Kiro Cave is represented by a red circle.

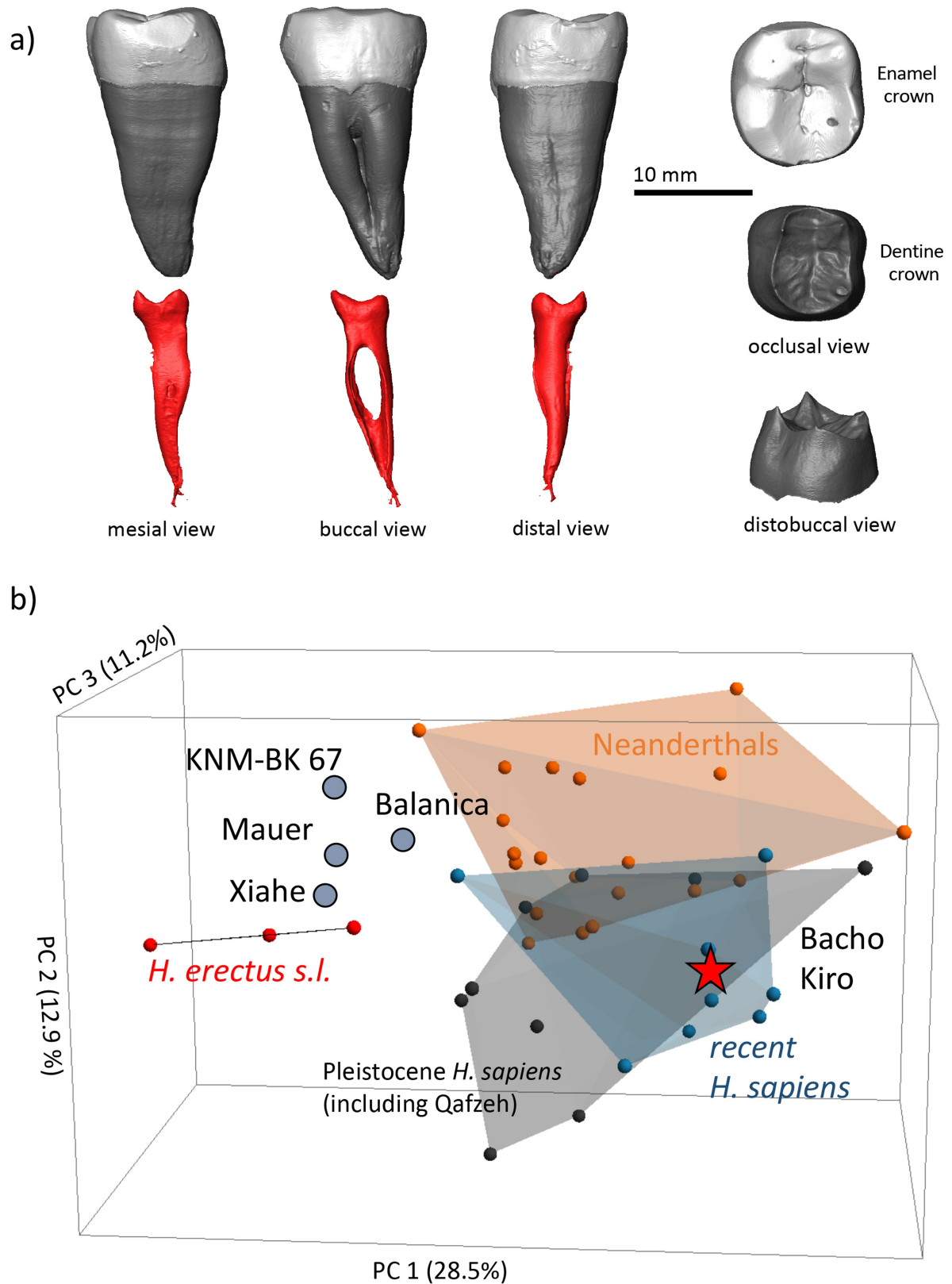


Extended Data Fig. 3 | Photographs of lithic artefacts from layer I of Bacho Kiro Cave. Pointed retouched blades and fragments (1–4, 6, 7) and piece with bifacial retouch (5). Photographs by V.S.-M. and T. Tsanova.



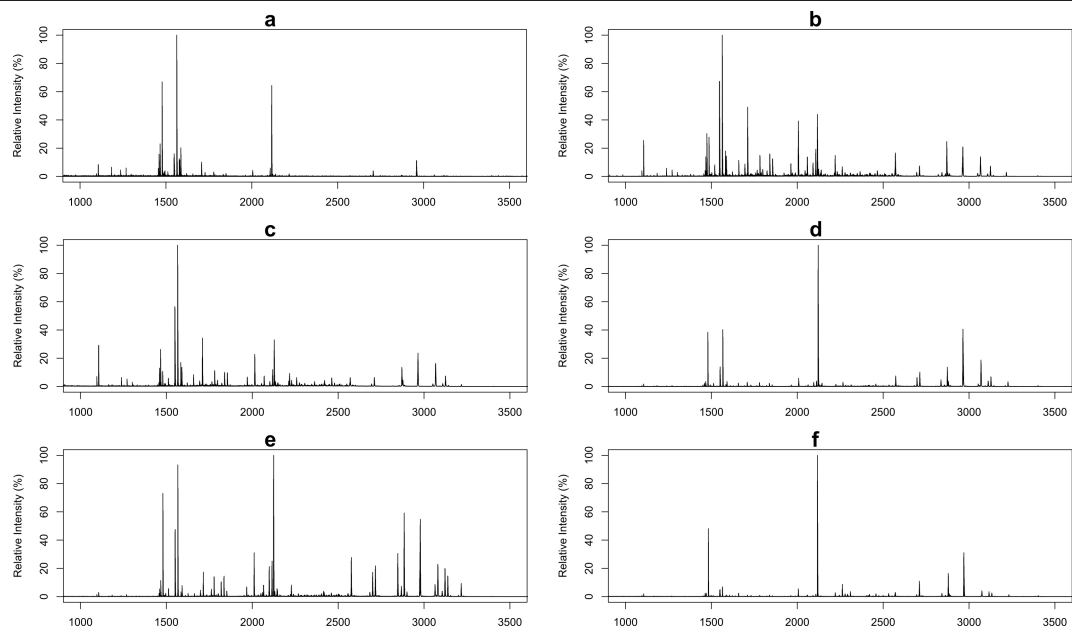
Extended Data Fig. 4 | Drawings of lithic artefacts from layer I of Bacho Kiro Cave. Pointed retouched blade with slightly oblique truncation and base modified by inverse retouch (1), pointed blade fragments (2 and 5, which has an oblique truncation and slight notch on the left edge, and was perhaps

intentionally fragmented), pointed, small blades fragments (3, 7, 8 and 9), pointed blade fragment with opposing pseudo-burin blows on the apex and on the distal fracture edge (perhaps indicating use as a projectile) (4) and Levallois flake (6). Drawings by I.K. and T. Tsanova).

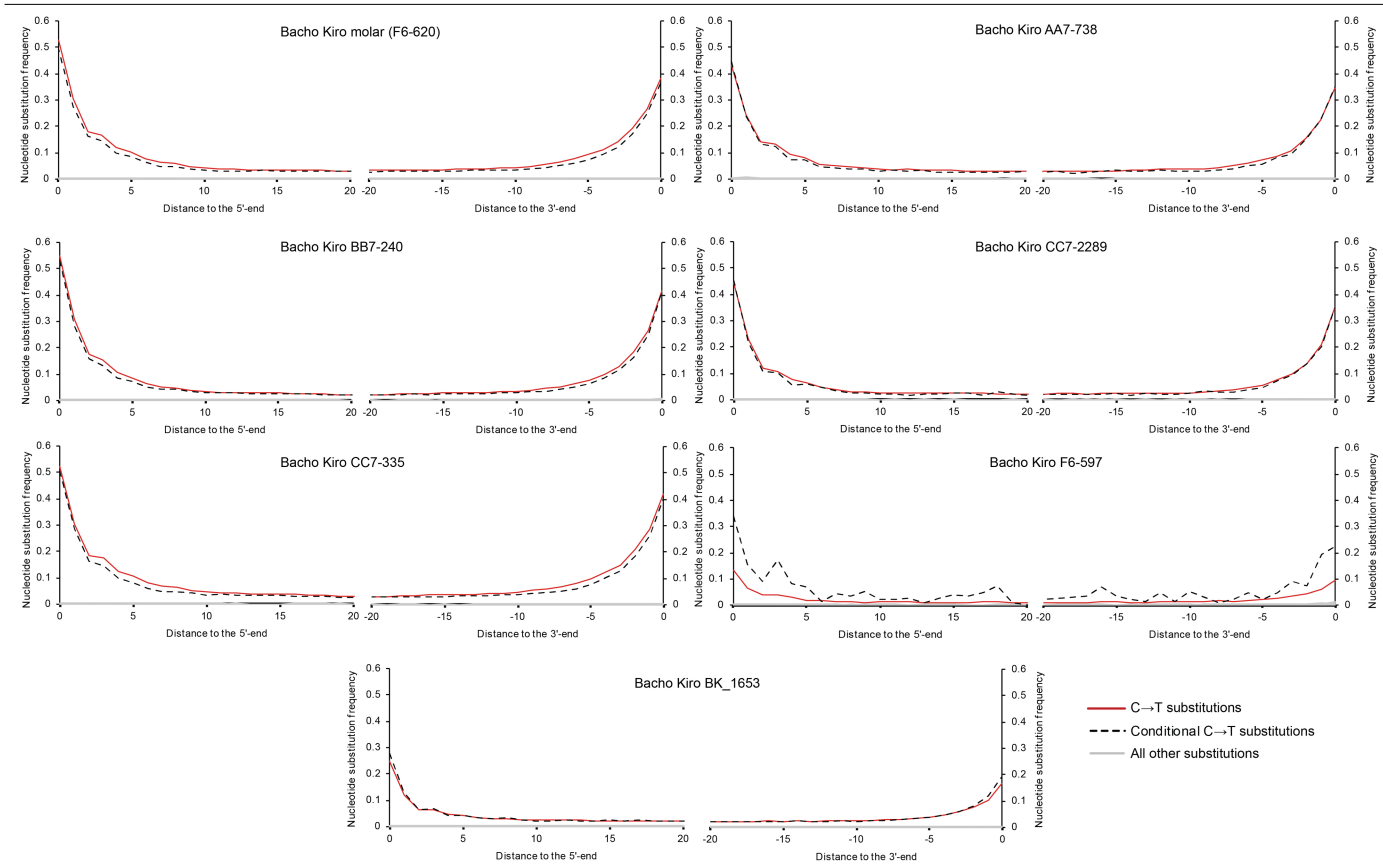


Extended Data Fig. 5 | Human lower second molar (F6-620). a, Mesial, buccal and distal views of the crown, root and pulp chamber (left) and occlusal views of the enamel and dentine crown (right). b, A principal component analysis of the shape of the enamel-dentine junction ridge and cervix places the Bacho

Kiro Cave second lower molar (F6-620) represented by a red star within the samples of recent ($n=8$) and Pleistocene ($n=9$) *H. sapiens*, and outside the distribution of Neanderthals ($n=20$) and *H. erectus* ($n=3$).

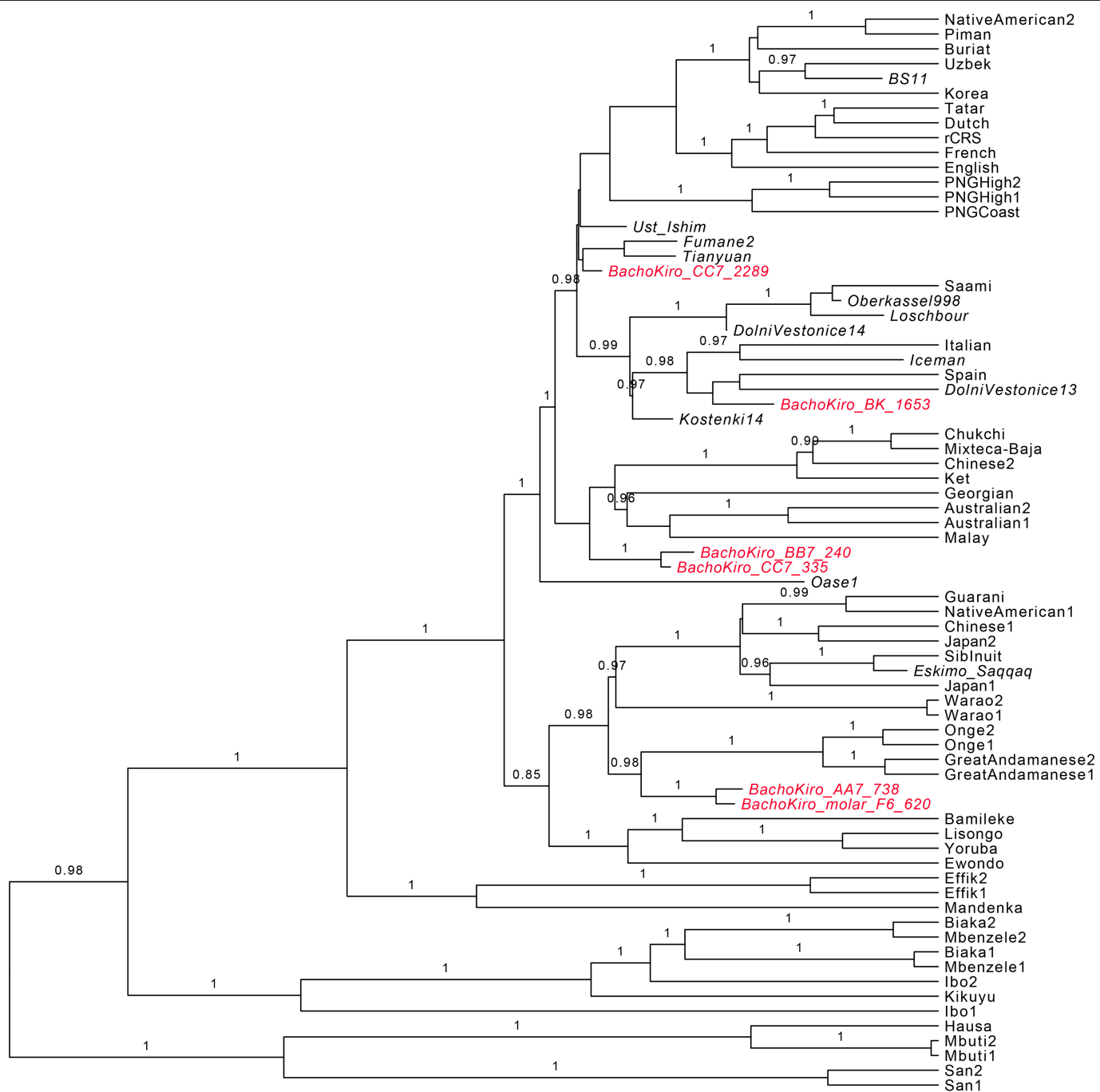


Extended Data Fig. 6 | MALDI-TOF MS spectra for the six bone specimens identified as hominins through ZooMS analysis. a, B4-1653 (interface of layers 6a and 7). b, AA7-738 (layer N1-I). c, BB7-240 (layer N1-I). d, CC7-2289 (layer N1-I). e, CC7-335 (layer N1-I). f, F6-597 (layer B).



Extended Data Fig. 7 | Frequency of nucleotide substitutions at the beginning and the ends of mtDNA alignments for the Bacho Kiro Cave specimens. Only fragments of at least 35 base pairs in length that mapped to the revised Cambridge Reference Sequence with a mapping quality of at least

25 were used for this analysis. Solid lines in red depict all fragments and dashed lines depict the fragments that have a C-to-T substitution at the opposing end ('conditional' C-to-T substitutions). All other types of substitution are marked in grey.



Extended Data Fig. 8 | Bayesian phylogenetic tree relating Bacho Kiro Cave mtDNA to 54 present-day humans, 10 directly radiocarbon dated ancient *H. sapiens* and the Vindija 33.16 Neanderthal. The Bacho Kiro Cave specimens are in red. Other ancient *H. sapiens* used as calibration points to

estimate the tip dates of Bacho Kiro Cave specimens are italicized. The posterior probabilities are denoted above the branches. The mtDNA of Vindija 33.16 was used to root the tree (not shown).

Extended Data Table 1 | Comparative dental metrics

		Bacho Kiro Cave	Neanderthal	Early <i>H. sapiens</i>	Upper Palaeolithic <i>H. sapiens</i>	Recent <i>H.sapiens</i>
M ₂	BL	11.23	\bar{x} =10.8 [9.6-12.4] σ =0.9 n = 31	\bar{x} =11.0 [9.2-12.7] σ = 1.0 n =22	\bar{x} =10.9 [8.6-12.3] σ =0.7 n=39	\bar{x} = 10.4 [8.6-12.5] σ = 0.8 n = 207
	MD	12.5	\bar{x} =11.6 [10.5-14.0] σ =1.0 n = 29	\bar{x} =11.7 [10.2-14.2] σ =1.1 n = 15	\bar{x} =11.3 [9.5-13.2] σ =0.8 n = 31	\bar{x} = 10.9 [8.9-14.3] σ = 0.9 n = 198
	CI	.89	.93	.95	.95	.96
	CCA	140.4	134.4	139.2	124.4	111.6

BL, bucco-lingual width; MD, mesiodistal length; CI, crown index (BL/MD); CCA, calculated crown area (BL × MD). Values are in mm. \bar{x} is the mean; minimum and maximum values are between the brackets; σ is the standard deviation; *n* indicates sample size. The Upper Palaeolithic *H. sapiens* sample includes individuals from the sites of: Les Abeilles, Bacho Kiro Cave, Brno, Bruniquel, Castenet, La Chaud, Dolní Věstonice, Farincourt, La Ferrassie, La Grèze, Les Rois, Isturitz, Kostenki, Kumchon, Laugerie-Basse, Lespugue, La Linde, Abri de la Madeleine, Nazlet Khater, Peștera cu Oase, Peche de la Boissiere, San Teodoro, St Germaine-la-Rivière, Sunghir, Les Vachons and Vindija. The early *H. sapiens* sample includes individuals from the sites of Border Cave, El Harhoura, Cave of Hearths, Dar es Soltane, Die Kelders, Haua Fteah, Jebel Irhoud, Klaises River Mouth, Mumba, Qafzeh, Skhul, Témara and Zhiren. The Neanderthal sample includes individuals from the sites of: Arcy-sur-Cure, Krapina, La Fate, Grotta Guattari, Hortus, Monte Fernera, Montmaurin, Ochoz, Petit-Puymoyen, La Quina, Le Regourdou, Spy, St Césaire, Subalyuk and Tabun. The recent human sample includes archaeological specimens representing western Europe, eastern Europe, southern Europe, Japan, China, the Near East, India, the Andaman Islands, Australia, New Guinea, northern Africa, southern Africa, eastern Africa and western Africa.

Extended Data Table 2 | mtDNA branch-shortening estimates

Specimen	Tip date estimate (years before present)	95% high posterior density intervals (HPD)
Molar (F6-620)	44,799	32,208-56,792
AA7-738	44,830	32,534-57,017
BB7-240	42,616	29,946-54,246
CC7-2289	44,532	30,203-56,128
CC7-335	42,670	30,085-54,243
BK-1653	30,763	20,602-39,544

Estimates for Bacho Kiro Cave specimens as determined in a Bayesian framework implemented in BEAST2, and by using 10 radiocarbon-dated ancient *H. sapiens* as calibration points (Supplementary Table 9).

Reporting Summary

Nature Research wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Research policies, see [Authors & Referees](#) and the [Editorial Policy Checklist](#).

Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

- | | |
|-------------------------------------|--|
| n/a | Confirmed |
| <input type="checkbox"/> | <input checked="" type="checkbox"/> The exact sample size (<i>n</i>) for each experimental group/condition, given as a discrete number and unit of measurement |
| <input checked="" type="checkbox"/> | <input type="checkbox"/> A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly |
| <input checked="" type="checkbox"/> | <input type="checkbox"/> The statistical test(s) used AND whether they are one- or two-sided
<i>Only common tests should be described solely by name; describe more complex techniques in the Methods section.</i> |
| <input checked="" type="checkbox"/> | <input type="checkbox"/> A description of all covariates tested |
| <input checked="" type="checkbox"/> | <input type="checkbox"/> A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons |
| <input type="checkbox"/> | <input checked="" type="checkbox"/> A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals) |
| <input checked="" type="checkbox"/> | <input type="checkbox"/> For null hypothesis testing, the test statistic (e.g. <i>F</i> , <i>t</i> , <i>r</i>) with confidence intervals, effect sizes, degrees of freedom and <i>P</i> value noted
<i>Give P values as exact values whenever suitable.</i> |
| <input type="checkbox"/> | <input checked="" type="checkbox"/> For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings |
| <input checked="" type="checkbox"/> | <input type="checkbox"/> For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes |
| <input checked="" type="checkbox"/> | <input type="checkbox"/> Estimates of effect sizes (e.g. Cohen's <i>d</i> , Pearson's <i>r</i>), indicating how they were calculated |

Our web collection on [statistics for biologists](#) contains articles on many of the points above.

Software and code

Policy information about [availability of computer code](#)

Data collection	Avizo 6.3; Geomagic Studio 2015.1.3; E4 software (http://www.oldstoneage.com/software/e4.shtml) and Microsoft Access 2016;
Data analysis	BATS 4.0; OxCal V4.3; Wolfram Mathematica 10; bam-rmdup 0.6.3; SAMtools 1.3.1; MAFFT v7.271; MEGA7, jModelTest2, mMass 5.5.0, R 3.4.2 and R 3.6.2, R package ggplot 2 3.2.1, R, package hexbin 1.28.0, RStudio 1.2.5033.

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors/reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Research [guidelines for submitting code & software](#) for further information.

Data

Policy information about [availability of data](#)

All manuscripts must include a [data availability statement](#). This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A list of figures that have associated raw data
- A description of any restrictions on data availability

Sequence reads from all libraries and corresponding negative controls are deposited at ENA under the study accession number PRJEB35466.

The FASTA files of the mitochondrial genomes are deposited in GenBank with the accession numbers MN706602-MN706607.

Details are as follows:

BachoKiro_AA7_738	MN706602
BachoKiro_BB7_240	MN706603
BachoKiro_BK_1653	MN706604
BachoKiro_CC7_335	MN706605
BachoKiro_CC7_2289	MN706606

BachoKiro_molar_F6_620 MN706607

Other data that support the findings of this study are available from the corresponding authors upon reasonable request

Field-specific reporting

Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

☐ Life sciences ☐ Behavioural & social sciences ☒ Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see [nature.com/documents/nr-reporting-summary-flat.pdf](https://www.nature.com/documents/nr-reporting-summary-flat.pdf)

Ecological, evolutionary & environmental sciences study design

All studies must disclose on these points even when the disclosure is negative.

Study description	Analysis of the palaeontological and archaeological material discovered in the course of a new excavation at the site of Bacho Kiro Cave (Bulgaria).
Research sample	The archaeological material (artifacts and faunal remains) was extracted from two areas (Main Sector and Niche 1) in the Bacho Kiro cave. Sediment samples were collected from each layer to perform micromorphological analyses. Collagen was extracted from archaeological bones from all stratigraphic levels of Bacho Kiro Cave to perform ZooMS and radiocarbon analyses.
Sampling strategy	For ZooMS, a random selection of morphologically unidentifiable bone specimens with a maximum length of over 2 centimeters was conducted. For radiocarbon dating and ancient DNA analyses, all human remains identified by ZooMS were sampled. For the morphological study of the human tooth F6-620, all comparative samples are detailed in the methods section.
Data collection	M.H., T.T., N.S., VA, S.S., R.S. Z.R. and S.P.M. collected field data; S.B., M.S. and J.J.H. collected morphological data on the hominin remains; V.S.M., L.P., F.W. and A.W. collected bone samples to perform ZooMS; M.H. extracted mtDNA from hominin bone samples; H.F. and S.T. collected bone samples to perform radiocarbon dating of hominin remains; G.S., R.S., V.P., N.M. collected morphological data on the faunal assemblage using the faunal reference collection stored at the Bulgarian National Museum of Natural History was used to accurately identify species and skeletal element; T.T. and S.P.M. collected metrical data on the lithic assemblage; V.A. collected micromorphological and sedimentological data in the site.
Timing and spatial scale	Bones were excavated from the Niche 1 and Main Sector areas of Bacho Kiro Cave during the 2015/2016/2017 field seasons. Bone pretreatment, and ancient DNA analyses were carried out over the course of 2016-2018.
Data exclusions	For the EDJ analysis, fossil teeth that were highly worn were excluded. For Radiocarbon dating two AMS dating methods across three AMS labs were used to check reproducibility. 11 collagen extracts from different layers were dated with graphite targets on a MICADAS AMS at two labs (ETH-Zurich and MAMS). Results were in statistical agreement for 8 of the extracts. Dates from the two labs were outside 2 sigma for 3 of the oldest extracts (all >40,000 BP). These samples were excluded from further analysis in the companion paper by Fewlass et al. Collagen from two human bones was dated with graphite targets at ETH-Zurich and in replicate with the gas ion source of the Aix-MICADAS AMS at CEREGE. All measurements were in statistical agreement.
Reproducibility	The mitochondrial genome sequences of Bacho Kiro Cave hominin specimens are deposited in GenBank.
Randomization	ZooMS samples were randomly analyzed. For the EDJ study we cannot determine any reason to apply randomization.
Blinding	For the EDJ study, blinding would be inappropriate given the small sample sizes and the relatively simple inferences made from the results of the principal component analysis.
Did the study involve field work?	<input checked="" type="checkbox"/> Yes <input type="checkbox"/> No

Field work, collection and transport

Field conditions	Yearly excavation inside the Bacho Kiro Cave of about one month between 2015-2018
Location	Bacho Kiro Cave, near Dryanovo (Bulgaria)
Access and import/export	The archaeological material was studied in Bulgaria. Temporary exports of some items were organized between Bulgaria and Germany. Permit delivered by National Museum of Natural History (Sofia) Nr. 4CH30/04.01.19.
Disturbance	The samples were obtained through archaeological excavation of two sections in the cave. At the end of the project (summer 2020) measures will be taken to protect the stratigraphic profiles.

Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

Materials & experimental systems

n/a	Involved in the study
<input checked="" type="checkbox"/>	<input type="checkbox"/> Antibodies
<input checked="" type="checkbox"/>	<input type="checkbox"/> Eukaryotic cell lines
<input type="checkbox"/>	<input checked="" type="checkbox"/> Palaeontology
<input checked="" type="checkbox"/>	<input type="checkbox"/> Animals and other organisms
<input checked="" type="checkbox"/>	<input type="checkbox"/> Human research participants
<input checked="" type="checkbox"/>	<input type="checkbox"/> Clinical data

Methods

n/a	Involved in the study
<input checked="" type="checkbox"/>	<input type="checkbox"/> ChIP-seq
<input checked="" type="checkbox"/>	<input type="checkbox"/> Flow cytometry
<input checked="" type="checkbox"/>	<input type="checkbox"/> MRI-based neuroimaging

Palaeontology

Specimen provenance	Excavation of the Bacho Kiro Cave authorized by the Bulgarian Ministry of the culture, delivered by NAIM-BAS: Nr124/11.05 2015; Nr225/28.04.2016; Nr47/02.05.2017; Nr99/17.04.2018/ Nr120/2019.
Specimen deposition	The palaeontological material will be deposited at the National Museum of Natural History in Sofia and the lithic material at the History Museum of Dryanovo (Bulgaria)
Dating methods	Small aliquots (80-110 mg) of the six ZooMS identified hominin bone fragments were sampled to preserve as much material as possible for further analyses. Collagen was extracted using a technique based on a modified Longin collagen extraction protocol followed by an ultrafiltration step. The gelatinized collagen samples were then passed through an Eze Filter (Elkay labs, UK) to remove large particles (>80 µm) and separated by molecular weight with pre-cleaned Sartorius VivaSpin Turbo 15 ultrafilters (30 kD MWCO). The samples were freeze dried and the large molecular fraction (>30 kD) was graphitised using the Automated Graphitisation Equipment III and measured using the latest model of the MICADAS AMS in the Laboratory of Ion Beam Physics at ETH-Zurich (lab code: ETH). The dates were calibrated using the IntCal13 dataset in OxCal v4.3.

☒ Tick this box to confirm that the raw and calibrated dates are available in the paper or in Supplementary Information.

Construction of a human cell landscape at single-cell level

<https://doi.org/10.1038/s41586-020-2157-4>

Received: 23 October 2018

Accepted: 12 March 2020

Published online: 25 March 2020



Xiaoping Han^{1,2,19}✉, Ziming Zhou^{1,19}, Lijiang Fei^{1,19}, Huiyu Sun^{1,19}, Renying Wang^{1,19}, Yao Chen^{3,19}, Haide Chen^{1,4,19}, Jingjing Wang^{1,4,19}, Huanna Tang⁵, Wenhao Ge⁶, Yincong Zhou⁷, Fang Ye¹, Mengmeng Jiang¹, Junqing Wu¹, Yanyu Xiao¹, Xiaoning Jia⁸, Tingyue Zhang¹, Xiaojie Ma⁹, Qi Zhang¹⁰, Xueli Bai¹⁰, Shujing Lai¹, Chengxuan Yu¹, Lijun Zhu⁶, Rui Lin¹¹, Yuchi Gao¹², Min Wang¹³, Yiqing Wu³, Jianming Zhang¹⁴, Renya Zhan¹⁵, Saiyong Zhu⁹, Hailan Hu⁸, Changchun Wang¹⁶, Ming Chen⁷, He Huang^{2,17,18}, Tingbo Liang¹⁰, Jianghua Chen⁵, Weilin Wang⁶, Dan Zhang³ & Guoji Guo^{1,2,4,17,18}✉

Single-cell analysis is a valuable tool for dissecting cellular heterogeneity in complex systems¹. However, a comprehensive single-cell atlas has not been achieved for humans. Here we use single-cell mRNA sequencing to determine the cell-type composition of all major human organs and construct a scheme for the human cell landscape (HCL). We have uncovered a single-cell hierarchy for many tissues that have not been well characterized. We established a ‘single-cell HCL analysis’ pipeline that helps to define human cell identity. Finally, we performed a single-cell comparative analysis of landscapes from human and mouse to identify conserved genetic networks. We found that stem and progenitor cells exhibit strong transcriptomic stochasticity, whereas differentiated cells are more distinct. Our results provide a useful resource for the study of human biology.

Individual cells are fundamental units of life. Breakthroughs in single-cell mRNA sequencing have greatly enhanced our ability to identify the transcriptomes of individual types of cell^{2–5}. Using high-throughput barcoding strategies, it is now possible to profile thousands of single cells at the same time^{6,7}. These methods have allowed the mapping of cell atlases for whole organisms^{8–15}. For example, cell atlases for the mammalian system have been generated by analysing both fetal and adult mouse tissues^{16–19}. Despite extensive efforts in dissecting the cellular compositions of various human tissues^{20–32}, to our knowledge a comprehensive cell landscape for humans has not been achieved.

Constructing an HCL using microwell-seq

Microwell-seq is a cost-effective single-cell mRNA sequencing technology that offers advantages over existing technologies in doublet rate and cell-type compatibility¹⁶. Sequencing titration experiments and cross-platform comparison suggest that this method can robustly detect rare populations even at low sequencing depth (Supplementary Table 1, Extended Data Fig. 1a). Using microwell-seq, we embarked upon the creation of a basic landscape of major human cell types using

samples from a Chinese Han population. Donated tissues were perfused or washed and prepared as single-cell suspensions using specific protocols (Supplementary Table 1). Our analyses included samples of both fetal and adult tissue and covered 60 human tissue types (two to four replicates per tissue type in general; Extended Data Fig. 1b). We also analysed seven types of cell culture, including induced pluripotent stem (iPS) cells, embryoid body cells, haematopoietic cells derived from co-cultures of human H9 and mouse OP9 cells³³, and pancreatic beta cells derived from H9 cells using a seven-stage protocol³⁴ (Extended Data Fig. 1b). Single cells were processed using microwell-seq¹⁶ and sequenced at around 3,000 reads per cell; data were then processed using published pipelines³⁵ (Fig. 1a). Altogether, 702,968 single cells passed our quality control tests (Supplementary Table 1).

In a global view, the complete human tissue data set is grouped into 102 major clusters (Fig. 1b; Supplementary Table 2). Multiple tissues, including artery, trachea, pleura, omentum, oesophagus and fallopian tube, contributed to the defined adult stromal/mesenchymal cells, such as cluster 4 (C4), C18, C27, and C70. Other clusters with substantial multi-tissue contributions correspond to fetal stromal cells (C7, C10, C17, C19, C21, C64, and C72), endothelial cells (C8, C20, C29, and C66), macrophages (C2, C51, C69, and C78), and fetal epithelial cells (C1)

¹Center for Stem Cell and Regenerative Medicine, Zhejiang University School of Medicine, Hangzhou, China. ²Bone Marrow Transplantation Center, The First Affiliated Hospital, Zhejiang University School of Medicine, Hangzhou, China. ³Department of Reproductive Endocrinology, Women's Hospital, Zhejiang University School of Medicine, Hangzhou, China. ⁴Zhejiang Provincial Key Lab for Tissue Engineering and Regenerative Medicine, Dr. Li Dak Sum & Yip Yio Chin Center for Stem Cell and Regenerative Medicine, Hangzhou, China. ⁵Kidney Disease Center, The First Affiliated Hospital, Zhejiang University School of Medicine, Hangzhou, China. ⁶Department of Hepatobiliary and Pancreatic Surgery, The Second Affiliated Hospital, Zhejiang University School of Medicine, Hangzhou, China. ⁷College of Life Sciences, Zhejiang University, Hangzhou, China. ⁸Center for Neuroscience, Zhejiang University School of Medicine, Hangzhou, China. ⁹Life Sciences Institute, Zhejiang University, Hangzhou, China. ¹⁰Department of Hepatobiliary and Pancreatic Surgery, The First Affiliated Hospital, Zhejiang University School of Medicine, Hangzhou, China. ¹¹Hangzhou Repugene Technology, Hangzhou, China. ¹²Annoroad Gene Technology, Beijing, China. ¹³Veritas Genetics Asia, Hangzhou, China. ¹⁴Department of Neurosurgery, The Second Affiliated Hospital, Zhejiang University School of Medicine, Hangzhou, China. ¹⁵Department of Neurosurgery, The First Affiliated Hospital, Zhejiang University School of Medicine, Hangzhou, China. ¹⁶Institute of Cancer Research and Basic Medical Sciences of Chinese Academy of Sciences, Cancer Hospital of University of Chinese Academy of Sciences, Zhejiang Cancer Hospital, Hangzhou, China. ¹⁷Institute of Hematology, Zhejiang University, Hangzhou, China. ¹⁸Stem Cell Institute, Zhejiang University, Hangzhou, China. ¹⁹These authors contributed equally: Xiaoping Han, Ziming Zhou, Lijiang Fei, Huiyu Sun, Renying Wang, Yao Chen, Haide Chen, Jingjing Wang. ✉e-mail: xhan@zju.edu.cn; ggj@zju.edu.cn

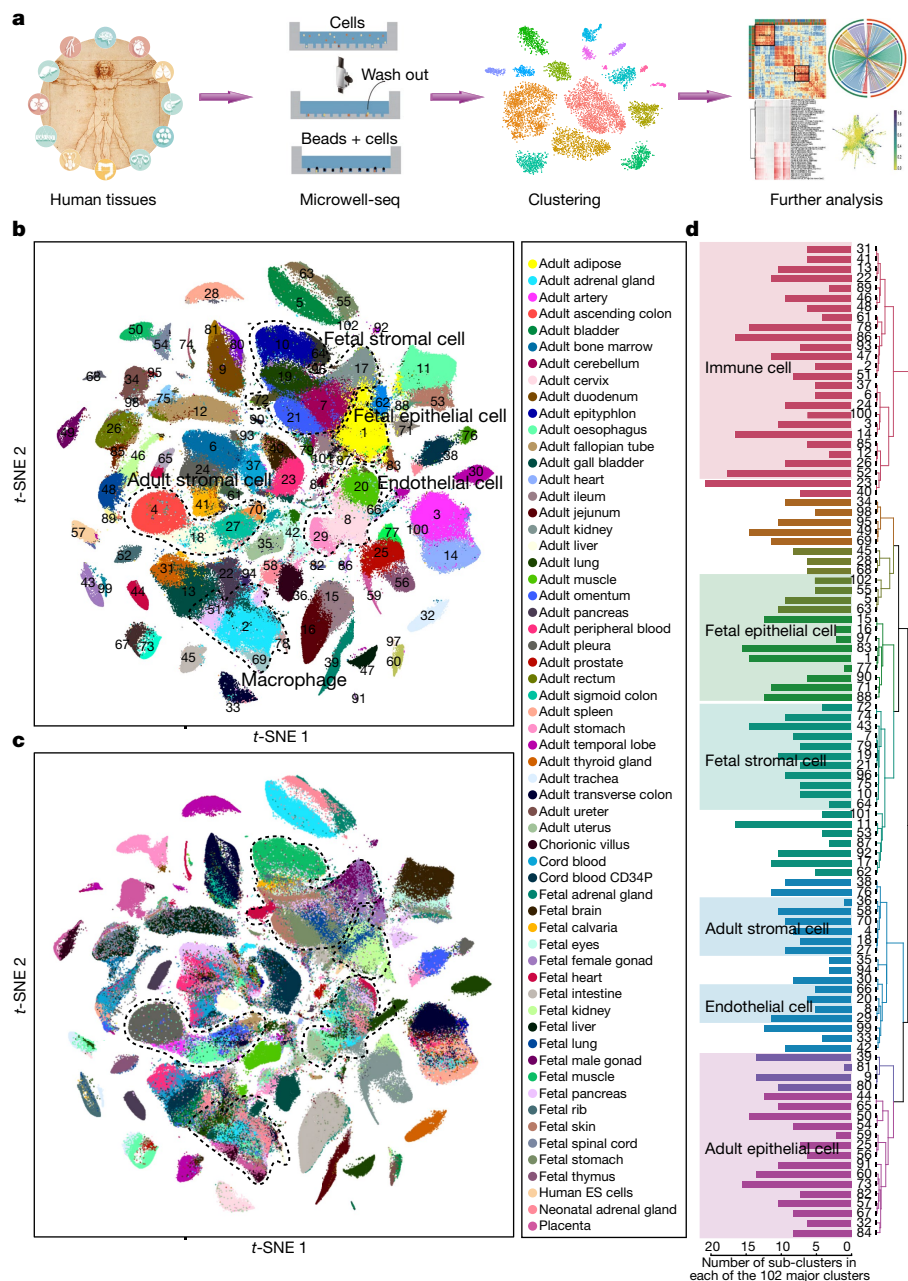


Fig. 1 | Constructing an HCL using microwell-seq. **a**, Illustration of the experimental workflow using the microwell-seq platform. **b**, *t*-SNE analysis of 599,926 single cells from the HCL. Differentiated cell culture data and granulocyte-colony stimulating factor (G-CSF)-mobilized peripheral blood data were not included. In the *t*-SNE map, 102 cell-type clusters are labelled by different colours. Cell cluster markers are listed in Supplementary Table 2. **c**, *t*-SNE analysis of 599,926 single cells from the HCL. Differentiated cell culture

data and G-CSF mobilised PB data were not included. In the *t*-SNE map, tissues are labelled by different colours. Tissue contributions to each cluster are listed in Supplementary Table 2. **d**, Dendrogram showing relationships among 102 cell types. The bar chart on the left represents the number of sub-clusters in each major cluster. A total of 843 sub-clusters were predicted from 102 major clusters.

(Fig. 1b, c; Supplementary Table 2). We then performed sub-clustering analysis for each of the 102 major clusters and predicted a total of 843 cell-type sub-clusters in the hierarchy (Fig. 1d). Through correlation analysis between bulk and single-cell mRNA sequencing as well as cell number sub-sampling analysis, we estimated a high gene and cell-type coverage of HCL (Extended Data Fig. 1c, d). Multi-donor analysis of representative tissues indicates that there are limited donor or batch effects on cell-type discovery (Extended Data Fig. 1e). To our knowledge, these data represent the most comprehensive cell-type repertoire yet described for the human species. By applying the concept of

a pseudo-cell³⁶, we can aggregate data from the same cluster to increase gene representation and improve cluster separation (Extended Data Fig. 2a, b). This strategy enabled us to interpret transcription factor (TF) function and generate a correlation network that covers 91% of all human TFs (Extended Data Fig. 2c; Supplementary Table 2). A highlight of the network suggests that, in the HCL, master TFs work in discrete modules to specify major human cell types such as neuron, erythroid cell, and acinar cell (Extended Data Fig. 2d). The resource is publicly available at <http://bis.zju.edu.cn/HCL/> (with a mirror website for international users at <https://db.cngb.org/HCL/>).

Cellular heterogeneity in human tissues

We performed *t*-distributed stochastic neighbour embedding (*t*-SNE) and differential gene expression analyses for each specific organ and uncovered previously unrecognized cell heterogeneity in a wide range of human tissues (Supplementary Table 3, Extended Data Figs. 3, 4). After analysing human kidney tissues at the fetal and adult stages, we defined 21 fetal clusters (FCs) and 22 adult clusters (ACs) with specific molecular markers (Extended Data Fig. 5a, b, Supplementary Table 3), including epithelial, endothelial, stromal and tissue-resident immune cells. We found previously undescribed types of S-shaped body cell (FC1, FC9 and FC15) in the fetal kidney and a new intercalated cell (IC)-tran-principal cell (PC) cell type (AC19) in the adult kidney. In the lung, we identified 22 fetal clusters and 25 adult clusters (Extended Data Fig. 5c, d, Supplementary Table 3). We found distal progenitors (FC4) and proximal progenitors (FC9) in the fetal lung, as well as alveolar type 2 (AT2) cells (AC1), alveolar type 1 (AT1) cells (AC3), two subsets of club cells (AC16 and AC20), and ciliated cells (AC22) in the adult lung. The alveolar bipotent/intermediate cells (AC14) express *KRT8*, cell cyclin genes, and lineage markers from both AT1 and AT2. We constructed ligand–receptor maps using CellPhoneDB³⁷ to reveal cell–cell interactions in kidney and lung tissues (Extended Data Fig. 5e–h). Stromal and endothelial cells were at the centre of the network for both organ types. In fetal tissues, our data predict that stromal and endothelial cells interact with epithelial cell progenitors to support tissue development. However, in adults, they are predicted to interact with immune cells such as T cells and macrophages.

During our comprehensive analysis, we noted the presence of populations with interesting expression signatures, particularly in tissues that have not previously been well characterized (summarized in Supplementary Table 4). In adult pleura, we identified three clusters of mesothelial cells (C2, C13 and C14), as well as an unknown cell cluster (C12) that expressed high levels of interferon-induced proteins (Extended Data Fig. 6a, b). In adult omentum, we identified a group of immune-related mesothelial cells that expressed *CCL2* (Extended Data Fig. 6c). In fetal muscle, we noted two groups of tendon cells expressing glucagon (*GCG*; Extended Data Fig. 6d). Other previously undescribed populations include *MGP*⁺ progenitor cells in fetal brain, stomach cells that co-expressed a D and X/A cell signature, and *CCL2*⁺ enteric nerve cells in the colon as well as 18 cell clusters that could not be annotated on the basis of known markers (Supplementary Table 4).

Notably, we identified endothelial cell populations that expressed major histocompatibility complex (MHC) class II genes such as *HLA-DPA1* and *HLA-DRA* in a variety of adult tissues, including bladder (Fig. 2a) and kidney. Immunofluorescence assays for the endothelial cell marker *PECAMI* (also known as *CD31*) and the MHC class II marker *HLA-DR* further confirmed the existence of antigen-presenting endothelial cells in the adult human bladder (Fig. 2b). This professional antigen-presenting signature indicates that these endothelial cells might have an immune function. To estimate the global involvement of non-immune cells in regional immunity in humans, we performed a cross-stage, cross-tissue comparison of endothelial, epithelial and stromal cells for their expression of immune-related genes (Fig. 2c). We identified MHC class II⁺ endothelial, interleukin-expressing stromal cells and *CXCL*⁺ epithelial cells in about half of human adult tissues (Extended Data Fig. 6e, f). This widespread immune activation of non-immune cells represents a new layer of cellular regulation for tissue-specific immunity.

To characterize the global cellular hierarchy of endothelial cells, we integrated endothelial cell data across diverse tissues and identified 14 major clusters (Fig. 2d, e, Supplementary Table 4). In the *t*-SNE map, C2, C3, and C9 contained contributions from multiple fetal tissues, including the heart, skin and kidney. Whereas C8, C11, C12 and C13 endothelial cells are tissue-specific, C6 and C10 are shared by different adult organs. Notably, there is a distinct group of endothelial cells,

including C1, C5 and C7, that show high expression of immune-related genes. C1 endothelial cells can be found in a wide range of adult organs such as bladder, kidney, artery, thyroid and omentum; they express *HLA-DRA*, *CD74* and *HLA-DRB1*. C5 endothelial cells come from adult uterus and express *CXCL8*, *IL6* and *HMOX1* at high levels. C7 endothelial cells come from adult kidney and express high levels of *HLA-DQA1*, *HLA-DPA1* and *EMCN*. In a similar way, we performed cross-tissue analysis for stromal cells and identified four immune-active clusters (C5, C8, C12 and C15) in the HCL stromal cell hierarchy (Extended Data Fig. 7a, b). Through donor contribution analysis, we showed that these stromal and endothelial subtypes contained cells from multiple donors (Extended Data Fig. 7c).

Fetal-to-adult cell-type transitions

We next studied human organ development by assessing the similarity of cell types between fetal and adult tissues. In the kidney and lung, the gene expression patterns of epithelial, mesenchymal, endothelial and immune cells were correlated between the two stages; tissue-resident immune and stromal cells appear early during organogenesis (Extended Data Fig. 8a, b). To visualize the process of global multi-lineage specification, we performed trajectory analysis for fetal and adult HCL data using the partition-based graph abstraction (PAGA) method³⁸. We obtained a landscape showing projections from fetal progenitors towards adult mature cell types (Extended Data Fig. 8c, d). On the landscape, we defined 36 cell groups arranged into more than 10 lineage branches. Notably, fetal cells reside at the centre of the landscape; they are less separated than adult cells, which spread out from the middle cloudy region towards different destinations. This is consistent with the observation in Fig. 1b that stromal and epithelial cells from various fetal tissues form an interconnected cell group (C1, C7, C17, C19 and C21). However, they become highly separated at the adult stage. Moreover, single cells exhibit higher entropy in the fetal stage than in the adult stage, suggesting that fetal cells possess higher transcriptional plasticity, whereas adult cells possess more stable transcriptomes (Extended Data Fig. 8e).

To understand the genetic regulation behind this fetal-to-adult transition, we analysed changes in gene expression between all specific pairs of fetal and adult cell types and ranked the most commonly enriched genes for the fetal and adult stages, respectively (Supplementary Table 4). Markers that were commonly enriched in fetal cells included a large group of ribosomal protein genes (such as *RPS17*, *RPS18*, and *RPS24*), imprinted genes (such as *IGF2* and *H19*), and stem cell regulators (such as *SOX4* and *SON*). These signatures suggest a high metabolic rate, low epigenetic restriction and the presence of multi-potent machinery, which may help to explain the high entropy and transcriptional stochasticity found in fetal cells. By contrast, markers that were commonly enriched in adult cells include MHC class I cell surface receptors (such as *HLA-A* and *HLA-B*), MHC class II cell surface receptors (such as *HLA-DPA1* and *HLA-DPBI*), and immune-related cytokines (such as *IFI27*, *CXCL2* and *CCL5*).

Mapping human cell types using scHCL

Using the HCL database, we build a single-cell mapping pipeline (scHCL) for the classification of human cell types. We integrated our HCL with other published human data sets (Supplementary Table 5) and made transcriptome references for all available human cell-type clusters from single-cell studies. Input digital gene expression (DGE) data were compared to each transcriptome reference to provide a match score based on gene expression correlation (Fig. 3a). By mapping bulk RNA sequencing data to our HCL reference, we can robustly define cell lineages of cultured cell populations³⁹ or cancer cell organoids⁴⁰ (Extended Data Fig. 9a, b). We then processed single-cell data from liver bud organoid cells⁴¹ and cerebral organoid cells⁴². While confirming the existence of

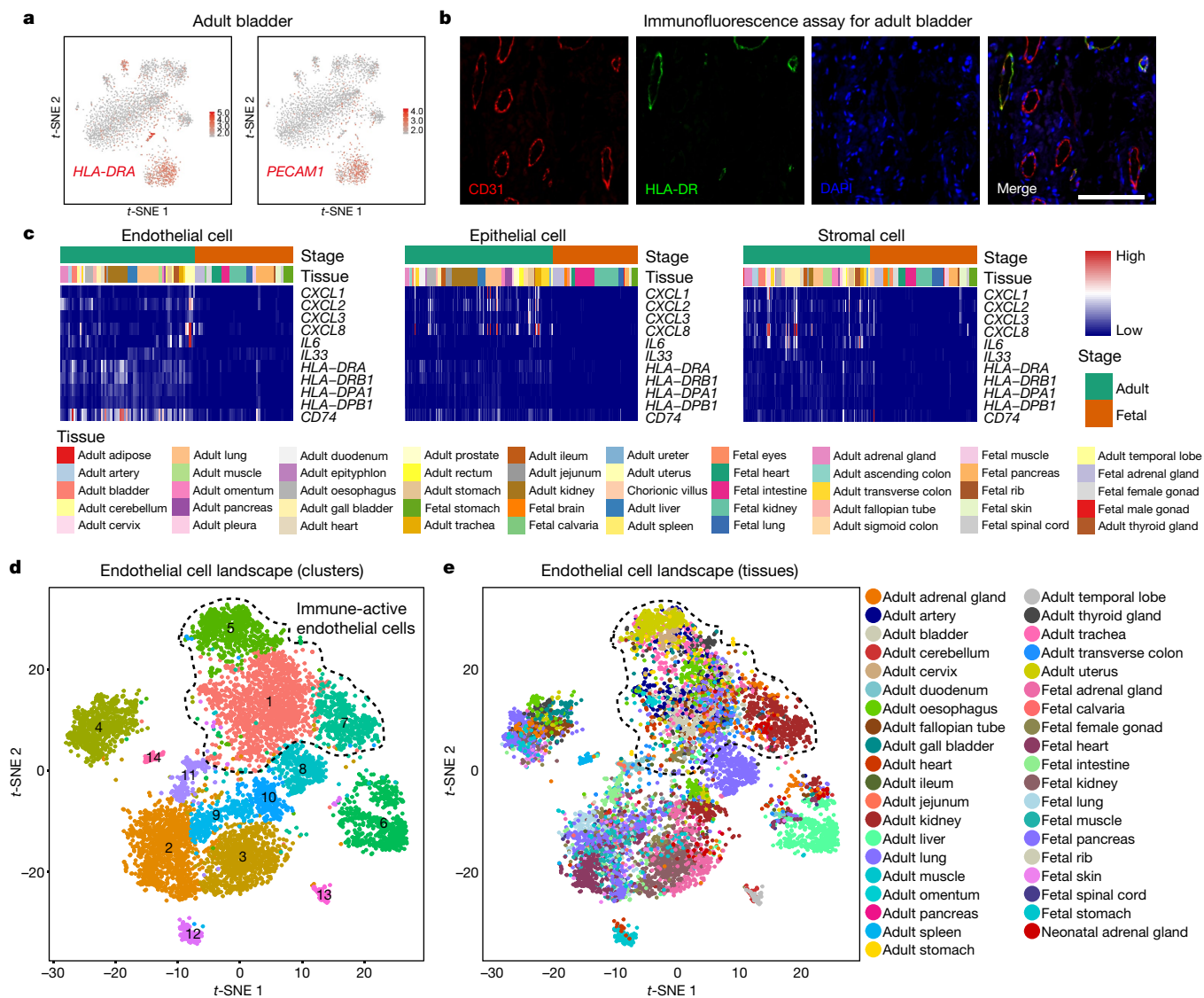


Fig. 2 | Immune activation of non-immune cells in the HCL. **a**, Feature plot in the *t*-SNE map of single-cell data from adult bladder cluster 2 ($n = 2,750$ cells). Cells are coloured based on the expression of *HLA-DRA* or *PECAM1*. The experiment was replicated twice with similar results. **b**, Immunofluorescence assay for the endothelial cell marker *PECAM1* (CD31) and the MHC class II marker *HLA-DR* in human adult bladder tissue. Scale bar, 100 μ m. The

experiment was replicated three times with similar results. **c**, Heat maps illustrating the expression of immune-related genes in endothelial, epithelial and stromal cells. *x*-axis represents clusters; *y*-axis represents genes. **d**, *t*-SNE maps of single-cell data for human tissue-specific endothelial cells (cell number $n = 7,140$). Cells are coloured according to endothelial cell subtype (**d**) or tissue type (**e**).

different parenchymal and mesenchymal cells (Extended Data Fig. 9c, 9d), we found that these organoids lacked important cell types, such as tissue-resident immune cells and endothelial cells.

We performed an scHCL analysis to evaluate a well-established seven-stage protocol for deriving pancreatic beta cells from H9 cells³⁴. Many cells present in the differentiated culture on day 24 exhibited a strong correlation with pancreatic cell types (Fig. 3b). Cells of C3, C5 and C8 highly expressed *GCG*, *INS* and *SST*, respectively, representing a tendency towards differentiation into alpha, beta and delta cells. However, a large proportion of these cells co-expressed islet signatures that correspond to various islet cell types, suggesting a remaining problem related to cell-type maturation in the differentiation protocol. We then used scHCL to analyse haematopoietic cells derived from H9/OP9 co-cultures³³. Only cells from C7 and C11, which contained endothelial cells and erythroid cells, had high scHCL scores for mature human cell types (Fig. 3c). The few C7 and C11 cells show low conversion efficiency in the H9/OP9 co-culture system. We did not detect any cell that was correlated with human haematopoietic stem and progenitor cells.

We next examined single-cell data for day 20 embryoid bodies differentiated from human iPS cells. The single cells of the embryoid body were divided into 15 clusters (Extended Data Fig. 9e, Supplementary Table 3). The scHCL results for the embryoid body cells showed clusters mapped to mesenchymal, neuronal, ependymal and immune cells (Extended Data Fig. 9f). A large group of cells in the middle of the *t*-SNE map, including C2, C3, C4 and C9, could not be associated with a specific lineage in the scHCL analysis. Moreover, the cell–cell correlation network of embryoid body cells showed complex and unspecialized phenotypes for these populations (Extended Data Fig. 9g). For example, C3 (Undefined-2) co-expressed the endothelial cell-related gene *EDNRB* and the neuron-related gene *MPZ*. Trajectory and RNA velocity analysis suggested that the undefined cell clusters were at the root of the differentiation hierarchy, with high RNA turnover rate, whereas the defined cell types are at the endpoint of the trajectory and have relatively stable transcriptomes (Fig. 3d). After differentiation for 20 days, many embryoid body cells appear to be still in a primarily undetermined state, whereas some cells have gone through these high turnover states and

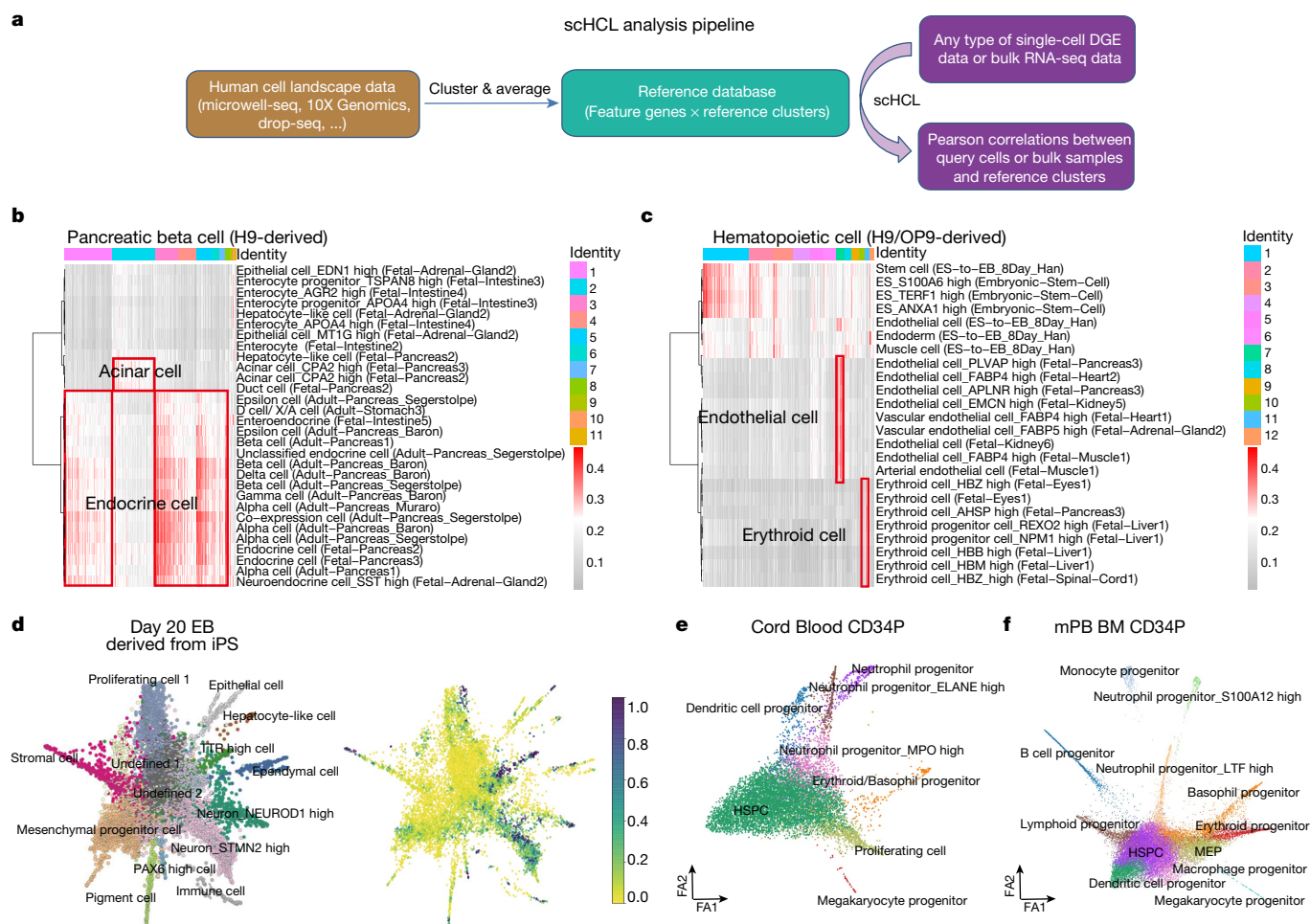


Fig. 3 | Application of scHCL analysis for stem cell biology. **a**, Diagram showing the pipeline for scHCL analysis. **b**, **c**, scHCL results for pancreatic islet cells (**b**; $n = 4,156$ cells) derived from seven-stage differentiation of H9 cells (day 24) and hematopoietic cells (**c**; $n = 1,115$ cells) derived from H9/OP9 cell co-cultures (day 9). Each row represents one cell type in our reference. Each column represents data from a single cell. Pearson correlation coefficient was used to evaluate cell-type gene expression similarity. Red indicates a high correlation; grey indicates a low correlation. Some cell-type data come from published works, as denoted by first author name: Segerstolpe²⁷, Muraro²⁸, Baron²⁹, Han⁵⁷. **d**, Branching gene expression trajectory analysis of day 20

embryoid bodies differentiated from human iPS cells. Left, trajectory of different lineages coloured by cluster identities; right, model of the transition probabilities derived from RNA velocity using a Markov process. The colour scale represents the density of the end points of the Markov process, from yellow (low) to blue (high). **e**, Branching gene expression trajectory analysis of cells in cord blood (CD34⁺) using PAGA, coloured by cell lineages. **f**, Branching gene expression trajectory analysis of cells in mPB (CD34⁺) using PAGA, coloured by cell lineages. Force atlas 1 (FA1) and Force atlas 2 are used to present a continuous graph layout.

exhibit homeostasis. This in vitro stem cell differentiation hierarchy mimics the in vivo fetal-to-adult cell-type landscape (Extended Data Fig. 8c, d). We also found similar patterns for the human haematopoietic differentiation system using single-cell data from CD34⁺ cord blood and CD34⁺ mobilized peripheral blood (mPB; Fig. 3e, f).

Comparison of mammalian cell landscapes

Single-cell transcriptomics offers the opportunity compare cell types across species. To compare the HCL and the mouse cell landscape, we re-clustered the updated data sets from the Mouse Cell Atlas (MCA), removed batch gene background, and performed analysis using the same pipeline and parameters as for the HCL (Extended Data Fig. 10a, b). To attenuate the effects of noise and outliers, we calculated pseudo-cells for each cell cluster to proceed with network construction (Extended Data Fig. 10c). Orthologous genes were extracted from the data to enable cross-species analysis. A correlation heat map among 102 human and 104 mouse cell types showed that cell-type similarity in orthologous gene expression overrides species differences,

particularly for immune and epithelial cells (Fig. 4a). As revealed by the circos plot, the gene expression patterns of major mammalian cell types are conserved; about 95% of cell types have a strong correlation (area under the receiver operating characteristics (AUROC) score > 0.9) between species (Fig. 4b). Tissue-specific cell-type pair analyses also suggested that the major cell types in mammalian organs are similar (Extended Data Fig. 10d, e). Our finding that mammalian cell types are conserved is consistent with the results of single-cell comparative genomics studies^{29,43,44}.

We next investigated the genetic network that underlies mammalian cell-type conservation. We extracted TF regulons from both HCL and MCA data using SCENIC⁴⁵ and identified 140 orthologous TF regulons grouped into 15 major modules across the mammalian cell landscapes (Fig. 4c, Supplementary Table 5). For example, M1, M6/8/11/13/15, M7 and M9 are associated with neuron, immune, endothelial and stromal cell types, respectively. These modules are enriched with not only lineage-specific transcription factors but also conserved binding motifs that lead to co-ordinated module activation in the lineage establishment (Fig. 4c).

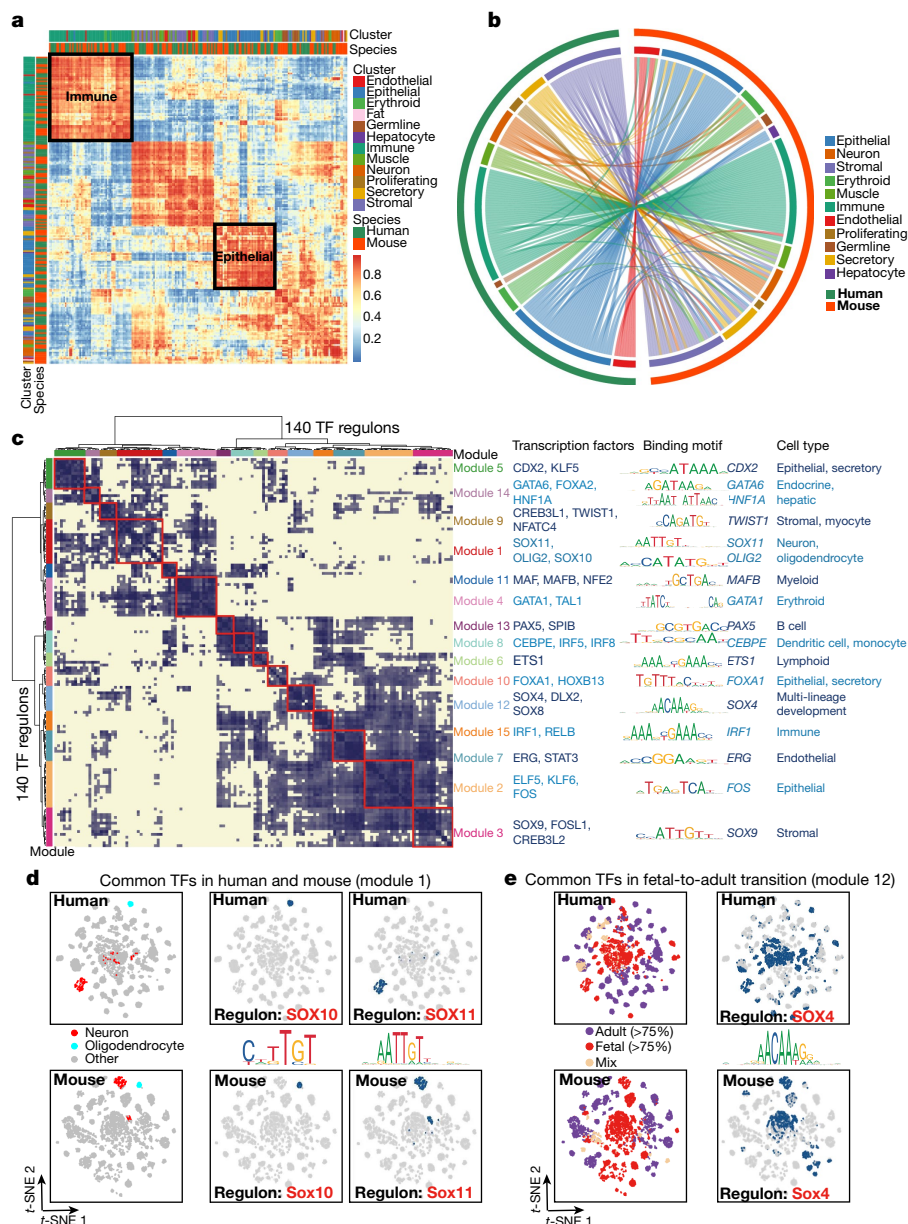


Fig. 4 | Cross-species comparison of cell landscapes. **a**, Correlation of orthologous gene expression between human and mouse cell types. AUROC scores were used to measure the similarity of cell types: red, high correlation; blue and yellow, low correlation, based on the Spearman correlation between all human and mouse pseudo-cells ($n = 46,793$ pseudo-cells). Cluster and species information is marked by different colours. **b**, Circos plot showing the similarity of cell types in human and mouse. Paired cell types with average AUROC scores greater than 0.9 are connected by lines. **c**, Identification of regulon modules based on the regulon matrix of the HCL and the MCA. The

network shows 140 orthologous TF regulons grouped into 15 major modules, along with representative TF regulons, corresponding binding motifs and associated cell types. **d**, **e**, Binary regulon activity scores (RASs) for regulons SOX10, SOX11 and SOX4 in the human and mouse regulon activity t-SNE map; dark blue dots represent one and grey dots represent zero. Other colours are used to mark specific cell types. **d**, Representative regulons in module 1. **e**, representative regulons in module 12. The regulon activity t-SNE maps were based on the binary regulon activity matrix of 17,028 human pseudo-cells and 16,740 mouse pseudo-cells.

To understand the specific genetic regulation of human and mouse cell types, we performed regulon activity analysis with all TFs (Extended Data Fig. 11a, b). We present a list of species-specific TF regulons enriched with basic helix–loop–helix, Cys2–His2 zinc finger and homeodomain proteins (Supplementary Table 5, Extended Data Fig. 11c–e). Nevertheless, most lineage-specific regulons are conserved (Supplementary Table 5, Extended Data Fig. 11f). For example, the SOX10 regulon for oligodendrocytes and SOX11 regulon for neurons are shared between human and mouse (Fig. 4d). Notably, the development-related SOX4 regulon dominates the unseparated fetal cell clusters; it shows a broad and stochastic distribution for both human and mouse (Fig. 4e).

Consistent with what we have seen for gene expression patterns, the stem and progenitor cell regulons lack lineage specificity and stability. However, regulons for differentiated cells appear to be more tightly wired; they may reach their steady state through continued self-reinforcement.

Discussion

We have used microwell-seq to perform single-cell transcriptomic analysis for a wide range of human tissues. The method proved to be compatible with nearly all cell types; microwell-seq data generated from different systems showed good comparability. The strength of

this landscape study is its broad coverage of both fetal and adult human tissue types. However, the scale of the current analysis is limited in sequencing depth and cell number for each individual tissue.

From the HCL, we identify MHC class II⁺ endothelial cells, CXCL⁺ epithelial cells, and interleukin-expressing stromal cells as major cell-type categories that lie between classical immune cells and non-immune cells. By comparing different developmental systems in the HCL, we propose a landscape model that is conserved in mammals: stem and progenitor cells are transcriptionally indistinct and stochastic; differentiated cells are transcriptionally distinct and stable; and the wiring of the regulons in the genome predetermines the terminal steady cellular states. By integrating different human single-cell data sets, we provide the HCL website and the associated scHCL pipeline for human cell-type identification. This pilot study will make contributions to the eventual completion of the international Human Cell Atlas.

Online content

Any methods, additional references, Nature Research reporting summaries, source data, extended data, supplementary information, acknowledgements, peer review information; details of author contributions and competing interests; and statements of data and code availability are available at <https://doi.org/10.1038/s41586-020-2157-4>.

1. Tanay, A. & Regev, A. Scaling single-cell genomics from phenomenology to mechanism. *Nature* **541**, 331–338 (2017).
2. Tang, F. et al. mRNA-seq whole-transcriptome analysis of a single cell. *Nat. Methods* **6**, 377–382 (2009).
3. Ramsköld, D. et al. Full-length mRNA-seq from single-cell levels of RNA and individual circulating tumor cells. *Nat. Biotechnol.* **30**, 777–782 (2012).
4. Treutlein, B. et al. Reconstructing lineage hierarchies of the distal lung epithelium using single-cell RNA-seq. *Nature* **509**, 371–375 (2014).
5. Shalek, A. K. et al. Single-cell transcriptomics reveals bimodality in expression and splicing in immune cells. *Nature* **498**, 236–240 (2013).
6. Macosko, E. Z. et al. Highly parallel genome-wide expression profiling of individual cells using nanoliter droplets. *Cell* **161**, 1202–1214 (2015).
7. Klein, A. M. et al. Droplet barcoding for single-cell transcriptomics applied to embryonic stem cells. *Cell* **161**, 1187–1201 (2015).
8. Plass, M. et al. Cell type atlas and lineage tree of a whole complex animal by single-cell transcriptomics. *Science* **360**, eaag1723 (2018).
9. Wagner, D. E. et al. Single-cell mapping of gene expression landscapes and lineage in the zebrafish embryo. *Science* **360**, 981–987 (2018).
10. Sebe-Pedros, A. et al. Cnidarian cell type diversity and regulation revealed by whole-organism single-cell RNA-seq. *Cell* **173**, 1520–1534.e1520 (2018).
11. Plasschaert, L. W. et al. A single-cell atlas of the airway epithelium reveals the CFTR-rich pulmonary ionocyte. *Nature* **560**, 377–381 (2018).
12. Fincher, C. T., Wurtzel, O., de Hoog, T., Kravarik, K. M. & Reddien, P. W. Cell type transcriptome atlas for the planarian *Schmidtea mediterranea*. *Science* **360**, eaag1736 (2018).
13. Farrell, J. A. et al. Single-cell reconstruction of developmental trajectories during zebrafish embryogenesis. *Science* **360**, eaar3131 (2018).
14. Cao, J. et al. Comprehensive single-cell transcriptional profiling of a multicellular organism. *Science* **357**, 661–667 (2017).
15. Spanjaard, B. et al. Simultaneous lineage tracing and cell-type identification using CRISPR-Cas9-induced genetic scars. *Nat. Biotechnol.* **36**, 469–473 (2018).
16. Han, X. et al. Mapping the mouse cell atlas by microwell-seq. *Cell* **172**, 1091–1107.e1017 (2018).
17. Tabula Muris Consortium. Single-cell transcriptomics of 20 mouse organs creates a Tabula Muris. *Nature* **562**, 367–372 (2018).
18. Cao, J. et al. The single-cell transcriptional landscape of mammalian organogenesis. *Nature* **566**, 496–502 (2019).
19. Pijuan-Sala, B. et al. A single-cell molecular map of mouse gastrulation and early organogenesis. *Nature* **566**, 490–495 (2019).
20. Zhong, S. et al. A single-cell RNA-seq survey of the developmental landscape of the human prefrontal cortex. *Nature* **555**, 524–528 (2018).
21. Young, M. D. et al. Single-cell transcriptomes from human kidneys reveal the cellular identity of renal tumors. *Science* **361**, 594–599 (2018).
22. Lake, B. B. et al. Integrative single-cell analysis of transcriptional and epigenetic states in the human adult brain. *Nat. Biotechnol.* **36**, 70–80 (2018).
23. Gao, S. et al. Tracing the temporal-spatial transcriptome landscapes of the human fetal digestive tract using single-cell RNA-sequencing. *Nat. Cell Biol.* **20**, 721–734 (2018).
24. Velten, L. et al. Human haematopoietic stem cell lineage commitment is a continuous process. *Nat. Cell Biol.* **19**, 271–281 (2017).
25. Li, L. et al. Single-cell RNA-seq analysis maps development of human germline cells and gonadal niche interactions. *Cell Stem Cell* **20**, 858–873.e854 (2017).
26. Guo, J. et al. Chromatin and single-cell rna-seq profiling reveal dynamic signaling and metabolic transitions during human spermatogonial stem cell development. *Cell Stem Cell* **21**, 533–546.e536 (2017).
27. Segerstolpe, Å. et al. Single-cell transcriptome profiling of human pancreatic islets in health and type 2 diabetes. *Cell Metab.* **24**, 593–607 (2016).
28. Muraro, M. J. et al. A single-cell transcriptome atlas of the human pancreas. *Cell Syst.* **3**, 385–394.e383 (2016).
29. Baron, M. et al. A single-cell transcriptomic map of the human and mouse pancreas reveals inter- and intra-cell population structure. *Cell Syst.* **3**, 346–360.e344 (2016).
30. Blakeley, P. et al. Defining the three cell lineages of the human blastocyst by single-cell RNA-seq. *Development* **142**, 3151–3165 (2015).
31. Yan, L. et al. Single-cell RNA-Seq profiling of human preimplantation embryos and embryonic stem cells. *Nat. Struct. Mol. Biol.* **20**, 1131–1139 (2013).
32. Aizarani, N. et al. A human liver cell atlas reveals heterogeneity and epithelial progenitors. *Nature* **572**, 199–204 (2019).
33. Vodyanik, M. A., Bork, J. A., Thomson, J. A. & Slukvin, I. I. Human embryonic stem cell-derived CD34⁺ cells: efficient production in the coculture with OP9 stromal cells and analysis of lymphohematopoietic potential. *Blood* **105**, 617–626 (2005).
34. Rezaania, A. et al. Reversal of diabetes with insulin-producing cells derived in vitro from human pluripotent stem cells. *Nat. Biotechnol.* **32**, 1121–1133 (2014).
35. Satija, R., Farrell, J. A., Gennert, D., Schier, A. F. & Regev, A. Spatial reconstruction of single-cell gene expression data. *Nat. Biotechnol.* **33**, 495–502 (2015).
36. Tosches, M. A. et al. Evolution of pallium, hippocampus, and cortical cell types revealed by single-cell transcriptomics in reptiles. *Science* **360**, 881–888 (2018).
37. Vento-Tormo, R. et al. Single-cell reconstruction of the early maternal-fetal interface in humans. *Nature* **563**, 347–353 (2018).
38. Wolf, F. A. et al. PAGA: graph abstraction reconciles clustering with trajectory inference through a topology preserving map of single cells. *Genome Biol.* **20**, 59 (2019).
39. Chu, L. F. et al. Single-cell RNA-seq reveals novel regulators of human embryonic stem cell differentiation to definitive endoderm. *Genome Biol.* **17**, 173 (2016).
40. Roerink, S. F. et al. Intra-tumour diversification in colorectal cancer at the single-cell level. *Nature* **556**, 457–462 (2018).
41. Camp, J. G. et al. Multilineage communication regulates human liver bud development from pluripotency. *Nature* **546**, 533–538 (2017).
42. Camp, J. G. et al. Human cerebral organoids recapitulate gene expression programs of fetal neocortex development. *Proc. Natl Acad. Sci. USA* **112**, 15672–15677 (2015).
43. La Manno, G. et al. Molecular diversity of midbrain development in mouse, human, and stem cells. *Cell* **167**, 566–580.e519 (2016).
44. Hodge, R. D. et al. Conserved cell types with divergent features in human versus mouse cortex. *Nature* **573**, 61–68 (2019).
45. Aibar, S. et al. SCENIC: single-cell regulatory network inference and clustering. *Nat. Methods* **14**, 1083–1086 (2017).

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

© The Author(s), under exclusive licence to Springer Nature Limited 2020

Methods

Ethics statement

The collection of human samples and research conducted in this study were approved by the Research Ethics Committee of the Zhejiang University School of Medicine, the Research Ethics Committee of the First Affiliated Hospital, the Research Ethics Committee of the Second Affiliated Hospital and the Research Ethics Committee of Women's Hospital at Zhejiang University (approval numbers: 20170029, 20180017, 20190034, 2018015, 2018507, 2018766 and 2018185). Informed consent for fetal tissue collection and research was obtained from each patient after her decision to legally terminate her pregnancy but before the abortive procedure was performed. Informed consent for collection of and research using surgically removed adult tissues was obtained from each patient before the operation. Informed consent for the collection and research of tissues from deceased organ donors was obtained from the donor family after the cardiac death of the donor. Details on donor information are provided in Supplementary Table 1. All the protocols used in this study were in strict compliance with the legal and ethical regulations of Zhejiang University School of Medicine and Affiliated Hospitals. All the protocols used in this study complied with the 'Interim Measures for the Administration of Human Genetic Resources' administered by The Ministry of Science and Technology and The Ministry of Public Health (approval number: 2020BAT0007). Mouse experiments in this study were approved by the Animal Ethics Committee of Zhejiang University; experiments conformed to the regulatory standards at Zhejiang University Laboratory Animal Center.

Fabrication of the microwell device

The diameter and depth of the microwells were 28 and 35 μm , respectively. First, a silicon plate containing 100,000 microwells was manufactured by Suzhou Research Materials Microtech Co., Ltd (Suzhou). The silicon microwell plate was then used as a mould to make a polydimethylsiloxane (PDMS) plate with the same number of micropillars. Before the experiments, a disposable agarose microwell plate was made by pouring 5% agarose solution onto the surface of the PDMS plate. Both the silicon and the PDMS plates are reusable. One silicon microwell plate allows almost permanent use.

Synthesis of barcoded beads

Magnetic beads (20–25 μm in diameter) coated with carboxyl groups were provided by Suzhou Knowledge & Benefit Sphere Tech. Co., Ltd (Suzhou; <http://www.kbspheretech.com/>). The barcoded oligonucleotides on the surface of the beads were synthesized by three rounds of split-pool. All the sequences used are the same as those reported previously¹⁶.

For each batch of bead synthesis, 300–350 μl carboxyl magnetic beads (50 mg/ml) were washed twice with 0.1 M 2-(*N*-morpholino) ethanesulfonic acid (MES). The beads were then suspended in a final volume of 635 μl of 0.1 M MES. 1-Ethyl-3-(3-dimethylaminopropyl) carbodiimide hydrochloride (EDC; 3.08 mg) was added to the beads, and 6.2 μl beads was then placed in each well of a 96-well plate. Amino-modified oligonucleotides (2.5 μl , 50 μM in 0.1 M MES) were then added to each well. After vortexing the mixture and incubating it for 20 min at ambient temperature, we distributed a 0.5- μl mix (6 mg EDC in 100 μl of 0.1 M MES) into each well. After an additional round of vortexing and incubation for 20 min at ambient temperature, an additional 0.5 μl mix (6 mg EDC in 100 μl of 0.1 M MES) was distributed into each well. After vortexing and incubation for 80 min at ambient temperature, the beads were collected in 1 ml of 0.1 M phosphate-buffered saline (PBS) containing 0.02% Tween 20. After centrifugation, the supernatant was carefully removed. The beads were then washed twice in 1 ml TE (10 mM Tris-HCl, 1 mM EDTA, pH 8.0).

In the second split-pool, the beads were washed with water and divided among the wells of another 96-well plate containing

polymerase chain reaction (PCR) mix (1 \times Phanta Master Mix, Vazyme) and 5 μM oligonucleotides. The oligonucleotides in each tube encoded a sequence with reverse complementarity to linker 1, a unique barcode and a linker 2 sequence. The PCR program was as follows: 94 $^{\circ}\text{C}$ for 5 min; five cycles of 94 $^{\circ}\text{C}$ for 15 s, 48.8 $^{\circ}\text{C}$ for 4 min, and 72 $^{\circ}\text{C}$ for 4 min; and a 4 $^{\circ}\text{C}$ hold. The third split-pool procedure was the same as the second one. The PCR program was as follows: 94 $^{\circ}\text{C}$ for 5 min, 48.8 $^{\circ}\text{C}$ for 20 min and 72 $^{\circ}\text{C}$ for 4 min, and a 4 $^{\circ}\text{C}$ hold. The beads were mixed sufficiently between denaturation (95 $^{\circ}\text{C}$) and primer annealing (48.8 $^{\circ}\text{C}$) in every cycle. The oligonucleotides used in each tube encoded a linker 2 reverse-complementary sequence, a unique barcode, a UMI sequence and a poly-T tail. All oligonucleotides were synthesized by Sangon Biotech Co., Ltd using high-performance liquid chromatography purification. To remove the chains without the third barcoded sequence, the beads were collected and suspended in 200 μl exonuclease I mix (containing 1 \times exonuclease I buffer and 1 U/ μl exonuclease I) and incubated at 37 $^{\circ}\text{C}$ for 15 min (the beads were mixed by a rotary mixer). After being washed with 200 μl TE-TW (10 mM Tris (pH 8.0), 1 mM EDTA, 0.01% Tween 20) and 200 μl of 10 mM Tris-HCl (pH 8.0), the beads were resuspended in 1 ml double-distilled water. To remove complementary chains, the beads were placed in a 95 $^{\circ}\text{C}$ water bath for 6 min and separated using a magnet, removing the supernatant quickly, twice. The beads could be stored in TE-TW for 4 weeks at 4 $^{\circ}\text{C}$. All oligonucleotide sequences are provided in Supplementary Table 1.

Mouse experiments to supplement the MCA database

Wild-type C57BL/6J mice were ordered from Shanghai SLAC Laboratory Animal. All mice were housed at Zhejiang University Laboratory Animal Center in a specific pathogen-free (SPF) facility with individually ventilated cages. The room has controlled temperature (20–22 $^{\circ}\text{C}$), humidity (30–70%) and light (12 h light–dark cycle). Mice were provided ad libitum access to a regular rodent chow diet.

Mouse adult adrenal gland, omentum, pleura and stomach tissues were collected from 8–10 week-old female mice. Mouse fetal pancreas and stomach tissues were collected from mouse embryos at embryonic day (E)14.5. Single-cell collections for these mouse tissues followed the same protocols as for the corresponding human tissues (Supplementary Table 1). Around 10,000 single cells were sampled for each tissue from 2–4 randomly chosen replicates. The data generated from these tissues were used to supplement the MCA database to enable human–mouse comparison.

Cell preparation

Samples from surgically removed adult tissues and aborted fetal tissues were put into Dulbecco's modified Eagle's medium plus 10% fetal bovine serum (FBS) immediately after resection. Samples from donations after cardiac death were resected after perfusion, and the time between cardiac death and tissue collection was controlled to within 1 h. Sample age (gestational age for fetal tissues), gender, cause of death, medical history were documented in Supplementary Table 1. Donated tissues were transported on ice from the hospital to the laboratory in less than 1 h, immediately transferred to cold Dulbecco's PBS (DPBS), and minced into pieces (~1 mm) on ice using scissors. The tissue pieces were transferred to a 15-ml centrifuge tube, rinsed twice with cold DPBS, and suspended in 5 ml solution containing dissociation enzymes. The samples were treated with various enzymes for different amounts of time (Supplementary Table 1). During dissociation, the tissue pieces were pipetted up and down gently several times until no tissue fragments were visible. The methods used for single-cell isolation from different tissues are listed in Supplementary Table 1. The dissociated cells were centrifuged at 300g for 5 min at 4 $^{\circ}\text{C}$ and then resuspended in 3 ml cold DPBS. After passage through a 40- μm strainer (Biologix), the cells were washed twice, centrifuged at 300g for 5 min at 4 $^{\circ}\text{C}$, and resuspended at a density of 1×10^5 cells/ml in cold DPBS containing

2 mM EDTA. CD34⁺ cord blood and CD34⁺ mPB were enriched using a human CD34 selection kit (StemCell Technologies).

Embryoid body differentiation

Human iPS cells (Sidansai) were cultured in hES cell medium with basic fibroblast growth factor (bFGF), dissociated using dispase and resuspended in embryoid body medium (hESC medium without bFGF). Cell clusters were seeded in six-well ultra-low adherent plates (Corning) at a density of 2×10^5 cells per well. The medium was changed every 3 days. On days 9, 18 and 20, embryoid body single cells were harvested by trypsinization and resuspended at a density of 1×10^5 cells/ml in cold DPBS containing 2 mM EDTA.

Cell collection and lysis

No fluorescence-activated cell sorting (FACS) step was used in any of the HCL experiments. All the analyses were single-cell analyses from fresh samples. Cell concentration was carefully controlled using a haemocytometer before microwell-seq. The proper cell concentration is $\sim 100,000$ /ml (with 10% of the wells occupied by single cells). The proper bead concentration is $\sim 1,000,000$ /ml (with every well occupied by single beads). An evenly distributed cell suspension was pipetted onto the microwell array, and extra cells were washed away. To eliminate cell doublets, the plate was inspected under a microscope. Cell doublets were reduced by pipetting over the region of high cell density. The bead suspension was then loaded into the microwell plate, and the plate was placed on a magnet. Excess beads were washed away slowly. Cold lysis buffer (0.1 M Tris-HCl (pH 7.5), 0.5 M LiCl, 1% sodium dodecyl sulfate (SDS), 10 mM EDTA, and 5 mM dithiothreitol) was pipetted over the surface of the plate and removed after 12 min incubation. The beads were then collected, transferred to an RNase-free tube, and washed once with 1 ml $6 \times$ SSC ($6 \times$ saline sodium citrate solution: 0.9 M sodium chloride and 0.09 M sodium citrate), once with 500 μ l $6 \times$ SSC, and once with 200 μ l of 50 mM Tris-HCl pH 8.0. Finally, $\sim 50,000$ beads were collected in a 1.5-ml tube.

Reverse transcription

In this procedure, the instructions from the Smart-seq2 protocol were followed. In brief, 20 μ l reverse transcription (RT) mix was added to the collected beads. The RT mix contained 200 U SuperScript II reverse transcriptase, $1 \times$ Superscript II first-strand buffer (Takara), 20 U RNase inhibitor (Sangon Biotech), 1 M betaine (Sigma), 6 mM MgCl₂ (Ambion), 2.5 mM dithiothreitol, 1 mM deoxynucleoside triphosphate, and 1 μ M TSO primer. The beads were incubated at 42 °C for 90 min with mixing on a rotary mixer and then washed with 200 μ l TE-SDS ($1 \times$ TE + 0.5% SDS) to inactivate reverse transcriptase. All oligonucleotide sequences are provided in Supplementary Table 1.

Exonuclease I treatment

The beads were washed with 200 μ l TE-TW and 200 μ l of 10 mM Tris-HCl (pH 8.0), resuspended in 100 μ l exonuclease I mix containing $1 \times$ exonuclease I buffer and 1 U/ μ l exonuclease I (NEB), and incubated at 37 °C for 60 min with mixing on a rotary mixer to remove oligonucleotides that did not capture mRNA. The beads were then pooled and washed once with TE-SDS, once with 1 ml TE-TW, and once with 200 μ l of 10 mM Tris-HCl (pH 8.0).

cDNA amplification

The beads were distributed into four PCR tubes. To each tube, 12.5 μ l PCR mix ($1 \times$ HiFi HotStart Readymix (Kapa Biosystems) and 0.1 μ M TSO_PCR primer) was added (Supplementary Table 1). The PCR program was as follows: 98 °C for 3 min; six cycles of 98 °C for 20 s, 65 °C for 45 s, and 72 °C for 6 min; 72 °C for 10 min; and a 4 °C hold. After all PCR products were pooled, AMPure XP beads (Beckman Coulter) were used to purify the cDNA samples from the PCR mix. Then, 25 μ l PCR mix ($1 \times$ HiFi HotStart Readymix and 0.1 μ M TSO_PCR primer) was

added to each DNA sample. The PCR program was as follows: 10 to 12 cycles of 98 °C for 3 min, 98 °C for 20 s, 67 °C for 20 s, and 72 °C for 6 min; 72 °C for 10 min; and a 4 °C hold. AMPure XP beads were used to purify the cDNA library.

Transposase fragmentation and selective PCR

The purified cDNA library was fragmented using a customized transposase that carries two identical insertion sequences. The customized transposase was included in the TruePrep DNA Library Prep Kit V2 for Illumina (Vazyme). The fragmentation reaction was performed according to the instructions provided by the manufacturer. We replaced the index 2 primers ($N5 \times \times$) in the kit with our P5 primer to specifically amplify fragments that contain the 3' ends of transcripts. Other fragments will form self-loops, impeding their binding to PCR primers. The PCR program was as follows: 72 °C for 3 min; 98 °C for 30 s; five cycles of 98 °C for 15 s, 60 °C for 30 s, and 72 °C for 3 min; 72 °C for 5 min; and a 4 °C hold. The PCR product was purified using AMPure XP beads. Then, 25 μ l PCR mix ($1 \times$ HiFi HotStart Readymix and 0.1 μ M 2100 primer) was added to each sample. The PCR program was as follows: 95 °C for 3 min; five cycles of 98 °C for 20 s, 60 °C for 15 s, and 72 °C for 15 s; 72 °C for 35 min; and a 4 °C hold. To eliminate primer dimers and large fragments, AMPure XP beads were then used to purify the cDNA library. The size distribution of the products was analysed on an Agilent 2100 bioanalyser, and a peak in the 400–700-bp range was observed. Finally, the samples were subjected to sequencing on the Illumina HiSeq systems. All oligonucleotide sequences are provided in Supplementary Table 1.

Immunohistochemistry

Donated human bladder and lung tissues were fixed in 4% paraformaldehyde overnight at 4 °C. Then, 30% sucrose/PBS was used to dehydrate the samples for 3 days at 4 °C. Immunohistochemistry was performed by Servicebio (Wuhan, China). In brief, the tissues were cut into 7- μ m-thick frozen sections and mounted on precleaned slides. The sections were washed three times with PBS and blocked with 5% FBS in PBS for 1 h at room temperature. Primary antibodies (anti-HLA-DR (1:200, EPR3692; Abcam), anti-CD31 (1:100, 66065-1-Ig; Proteintech), anti-Krt17 (1:50, MA5-13539; Thermo Fisher), and anti-CXCL2 (1:100, AHP773; Bio-rad)) diluted in blocking solution (5% FBS in PBS) were added to cover the sections. The slides were placed in a wet box and incubated at 4 °C. Relevant AlexaFluor488/594-conjugated secondary antibodies (1:1,000, Invitrogen) were used for labelling. The slides were then washed three times with blocking solution and stained with DAPI. Glass coverslips were then attached to the slides using mounting medium. Immunofluorescence images were obtained using confocal microscopy.

Processing of microwell-seq data

Microwell-seq data sets were processed as described¹⁶. Reads from HCL data were aligned to the *Homo sapiens* GRCh38 genome using STAR⁴⁶ and the DGE data matrices were obtained using the Dropseq core computational protocol (available at <http://mccarrolllab.org/dropseq/>) with default parameters. For quality control, we filtered out cells with detection of fewer than 500 transcripts. Cells with high proportion of transcript counts derived from mitochondria-encoded genes were also excluded.

Clustering of single-cell data matrix

Seurat³⁵ was used to perform clustering analysis of single-cell data from different tissues. DGE data were used as inputs. Cells from the pre-processed data and genes expressed in more than three cells were selected for further analysis. Filtered data were $\ln(\text{CPM}/100 + 1)$ transformed, and the number of UMI and the percentage of mitochondrial gene content were regressed out according to the published method⁴⁷. About 2,000 genes with an average expression of more than 0.01 and

a dispersion greater than 0.5 were used as inputs for initial principal component analysis (PCA) and the number of principal components (PCs) used for nonlinear dimensional reduction (*t*-SNE) analysis was chosen according to the PCElbowPlot function and JackStrawPlot function. For clustering, we set different resolution parameters between 0.6 and 4 in FindAllCluster function and narrowed down to certain cluster numbers by distinguishing differential genes among clusters. The heat map produced by DoHeatmap function is one basis for judging the quality of clustering. These parameters, including the resolution and number of PCs, were adjusted on per-tissue basis. The default Wilcoxon rank-sum test was used by running FindAllMarkers function in Seurat to find differentially expressed markers in each cluster. Finally, we annotate each cell type by extensive literature reading and searching for the specific gene expression pattern.

Batch removal for cross-tissue comparison

To improve the presentation, we strictly removed the batch gene background for cross-tissue comparison. We assumed that, for each batch of experiments, cell barcodes with fewer than 300 UMIs correspond to empty beads exposed to free RNA during cell lysis, RNA capture and washing steps. Genes with extensive expression in all beads were considered batch genes. The batch gene background value was defined as the average gene detection for all cellular barcodes with fewer than 300 UMIs multiplied by the median of the fold difference between the detected gene expression of a cell and the average detected gene expression for beads with fewer than 300 UMIs; this value was then rounded to the nearest integer. We subtracted the batch gene background for each cell from the digital expression matrix before performing the cross-tissue comparison. We used the background-removed matrix to perform cross-tissue analyses such as the cross-tissue comparison of endothelial cells and stromal cells over the whole body.

Landscape construction

For clustering of the complete human tissue data set in Fig. 1, we excluded data from differentiated cell cultures and G-CSF-mobilized PB to ensure natural status of the cell landscape. A total of 599,926 cells were selected and processed using Scanpy⁴⁸ in python environment. Background-removed DGE data with cells analysed in each tissue and genes expressed in at least 20 cells were used as inputs for Scanpy. Then, DGE data were $\ln(\text{CPM}/100 + 1)$ transformed. We selected about 3,000 highly variable genes according to their average expression and dispersion. We then regressed out UMI and gene numbers and scaled each gene to unit variance with clip values exceeding a standard deviation of 10. We chose 50 PCs for PCA and computed the neighbourhood graph of cells. We then used the Louvain clustering to cluster cells with resolution of 3.5 and $k = 15$. Finally, 102 clusters for the human landscape were produced and marker genes were calculated by the Wilcoxon rank-sum test. For sub-clustering of 102 clusters, we used the Seurat pipeline described previously with the default resolution of 0.8 to process cells from each cluster and predicted a total of 843 sub-clusters. The dendrogram in Fig. 1d was drawn using Pearson correlation coefficient (PCC) with mean gene expression for each cluster (CPM/100 values of all genes) with the R package denextend⁴⁹.

Pseudo-cell analysis

To increase the gene number and gene expression correlation from high-throughput single-cell mRNA data, we aggregated data from multiple cells in the same cell cluster to make pseudo-cells for genetic network interpretation⁵⁰. To justify the use of pseudo-cells, we compared the performance of different levels of pseudo-cells in three mouse lung data sets from three different technologies: microwell-seq, 10X Genomics and Smart-seq2. Gene number, gene expression correlation, and silhouette value were used to evaluate the performance (Extended Data Fig. 2a). A high silhouette value indicates a high degree of separation

among cell types using pseudo-cell data. The global *t*-SNE map based on pseudo-cells indicates improved separation among cell types (Extended Data Fig. 2b). On the basis of human pseudo-cell 20 data, we calculated the TF–TF correlation using PCC and generated a correlation heat map that covers 1,521 human TFs (Extended Data Fig. 2c). A correlation threshold of 0.5 was used to construct a high-confidence TF–TF correlation network (Extended Data Fig. 2d).

Receptor–ligand pairing analysis

Analysis of potential receptor–ligand pairings was performed by applying the recently published method CellPhoneDB³⁷. First, we aggregated the gene expression levels of 20 cells from each cluster in the adult lung, adult kidney, fetal lung and fetal kidney. To eliminate the effect of variable cell numbers in each cluster, we randomly sampled three pseudo-cells for analysis. Only receptors and ligands expressed in more than 10% of the cells in the specific cluster were considered. By permuting cluster labels randomly 1,000 times to calculate the mean expression values of ligands and receptors, interaction was constructed as a receptor–ligand pairing matrix. Then, we used pairwise comparisons between all cell types and obtained a likelihood of *P* value to filter the false-positive interaction. The cutoff was set with the mean expression greater than 0.1 and *P* values smaller than 0.1. We used the sum of the number of receptor–ligand pairs in each cell–cell pairing to indicate the strength of the cell–cell interactions. Finally, the network was set to degree sorted circle layout for visualization in Cytoscape⁵¹.

RNA velocity analysis

We used velocity⁵² to calculate RNA velocity on data from the human embryoid body on day 20. Velocity estimates the rate of transcriptional changes of each cell based on the ratio of spliced and unspliced reads. We used the ‘velocity run’ function described on the website (<http://velocityto.org/velocityto.py/tutorial/index.html>) with default parameters. Then, we imported the loom file from the last step and discarded cells that either did not have enough UMIs after the new mapping or did not have unspliced reads. Based on the coefficient of variation and average expression, we selected 2,013 genes to perform a PCA. Sixty-five PCs were used to impute with $k = 5$ nearest neighbours. The plot was visualized with *t*-SNE embedding and the differentiation start and end points were estimated using a Markov process. We mainly followed the steps from the open repository (https://github.com/rajewsky-lab/planarian_lineages).

Differential gene expression analysis between fetal and adult cell lineages

For differential gene expression analysis in Supplementary Table 4, we aggregated data from 20 cells in the same cluster to make pseudo-cells for each cell type. Then we used MetaNeighbour⁵³ to find the related cell-type pairs between fetal and adult tissues. In brief, MetaNeighbour was obtained through neighbour voting based on the Spearman correlation between all fetal and adult pseudo-cells. Then the mean AUROC scores were obtained from MetaNeighbour. Cell type pairs with AUROC score > 0.9 were regarded to have strong relationships in the development. Next, we performed the Wilcoxon rank-sum test to find the differential expressed genes within these cell type pairs. Genes with adjusted $P < 0.05$ for each cell-type pair were then labelled. Common top differential genes were estimated by the frequency of differential expression in all fetal-to-adult cell-type pairs.

Single-cell trajectory analysis

We used PAGA in scanpy⁴⁸ to infer the lineage tree of single-cell data from non-immune cell types in fetal and adult tissue as well as CD34⁺ haematopoietic data sets. The graph abstraction algorithm reconciles clustering and trajectory inference by explaining data variability in terms of both discrete and continuous latent variables. First, we

processed the data following the steps suggested by scanpy, including total count normalization, log1p logarithmization, highly variable genes extraction, a potential regression of confounding factors of genes and counts, a scaling to z-scores and PCA analysis. Then, we computed a neighbourhood graph among data points and used UMAP to generate a topologically faithful embedding with $\text{min_dist} = 0.1$. Then, PAGA was performed; with $\text{iter} = 1,000$, the trajectory was constructed using layout 'fa'. Differential gene expression analysis was performed with the `tl.rank_genes_groups()` function in scanpy with Wilcoxon rank-sum test.

Single-cell entropy analysis

We applied single cell lineage inference using cell expression similarity and entropy (SLICE) for quantitative measurements of cell differentiation states based on the calculation of single-cell entropy (scEntropy)⁵⁴. We performed deterministic calculation of scEntropy of individual cells in fetal and adult tissues with default parameters according to the SLICE pipeline. Similar results can also be obtained using the StemID method⁵⁵.

HCL website construction

The main HCL website uses a bootstrap framework to improve overall adaptability and interactivity. Its back-end is completed by PHP, R language and MySQL. The main functions of the HCL website are divided into four parts: gallery, landscape, search and schCL. Gallery provides interactive *t*-SNE maps for more than 80 data sets to show the distribution of different clusters. Specific markers for each cluster are listed in a data table. Landscape achieves better visualizations for global view of 102 clusters using both single cells and pseudo-cells. Search describes the expression of a given gene in different clusters from any selected tissue. schCL provides the function of single-cell correlation analysis with the HCL database. After users upload their own DGE files, the data are processed by the R script and compared to the HCL reference file. The schCL result is returned in JSON format and presented as an interactive heat map.

schCL analysis

Similar to our previously published single-cell mapping pipeline¹⁶, the schCL analysis was conducted through the following steps. We downloaded published drop-seq, in-drop, seq-well and 10X Genomics data from human samples^{11,20,22,24–31,37,41,43,56–62} (Supplementary Table 5). We redid the clustering for each data set, combined the cell-type clusters with our HCL clusters and generated a total of 1,841 cell-type clusters (cell-type redundancies exist in these clusters). Then, for each cell-type cluster, we randomly sampled 100 single cells without replacement (all cells for clusters with fewer than 100 cells), calculated the average expression normalized to 100,000 transcripts, and rounded the number down to the nearest integer. We constructed the averaged cell-type transcriptome data three times for each cell cluster. This resulted in main transcriptome references in our schCL pipeline. We then performed differential gene expression analysis for each cell type against all the other cell types and selected the top 20 marker genes for each cell type ($\log \text{foldchange} > 1$). Markers for each cell type were merged to create the combined feature gene list. The PCCs of the given single-cell data against each HCL cell-type reference were then calculated using the combined feature gene list. To properly display the mapping results of schCL, the number of top cell-type hits for each single cell (*nCluster*) and the lowest correlation coefficient threshold (*corThreshold*) can be manually adjusted. For all schCL results shown in this paper, *nCluster* was set to be 1; only the top 1 mapping hit was shown for each data point. The single-cell fragments per kilobase of transcript per million mapped reads (FPKM), reads per kilobase of transcript per million mapped reads (RPKM), transcripts per million (TPM) digital gene expression (DGE) and bulk RNA DGE matrix can also be applied to the schCL pipeline.

Cross-species transcriptome comparison

Before transcriptome comparison, we supplemented the MCA database with new microwell-seq data set from adult adrenal gland, omentum, pleura and stomach tissues, as well as fetal pancreas and stomach tissues. We followed the HCL pipeline to re-cluster the updated MCA data (https://figshare.com/articles/MCA_DGE_Data/5435866) ($n = 333,778$) into 104 major clusters (Extended Data Fig. 10a, b). We then divided the 102 HCL and 104 MCA cell clusters into 12 major cell lineages. To make the gene expression profiles of cross-species cell types comparable, we downloaded the homology correspondences between human and mouse provided by dmod-ENCODE⁶³. The gene expression profiles for human (this study) and mouse¹⁶ were normalized to the total number of transcripts and multiplied by 100,000. To attenuate the effects of noise and outliers, we used pseudo-cells³⁶ for further analysis; each pseudo-cell was an average of 20 cells randomly selected from the same cell type. To compare cross-species transcriptomes, we performed MetaNeighbour⁵³ analysis through neighbour voting based on the spearman correlation between all human and mouse pseudo-cells. Then the mean AUROC scores were obtained from MetaNeighbour. The Circlize⁶⁴ package was used to view the similarity of cell-type pairs between species, whose AUROC scores were higher than 0.9 or match types were 'Reciprocal_top_hit'.

Cross-species regulatory network comparison

Human and mouse gene regulatory network analysis was performed using SCENIC⁴⁵ with default parameters. In brief, the gene regulatory network was based on co-expression and DNA motif analyses, and then the AUCell algorithm was used to score the activity of each TF regulon in each pseudo-cell. From SCENIC analysis, we obtained a total of 259 human TF regulons and 248 mouse TF regulons, among which we defined 119 human unique TF regulons, 108 mouse unique TF regulons and 140 orthologous TF regulons (Supplementary Table 5). Based on 140 orthologous TF regulons, we obtained a merged AUCell score matrix from both human and mouse pseudo-cells. Then the PCC of AUCell score was calculated for each pair of orthologous TF regulons. From the analysis, we identified 15 orthologous TF regulon modules based on the connection specificity index (CSI)⁶⁵ using the 'ward.D' clustering method. For a fixed pair of regulons, a and b, the corresponding CSI was defined as the fraction of regulons whose PCC of AUCell scores with a and b were lower than the PCC between a and b themselves. The distribution of mean AUCell scores across the cell types was used to present the relationship between orthologous TF regulon modules and cell types.

Regulon activity analysis

To determine the 'on/off' activity of each regulon (259 TF regulons for human and 248 TF regulons for mouse) in each cell type, we used ' $0.5 * \max(\text{AUC scores})$ ' for each regulon as a threshold to binarize the regulon activity scores and created the 'binary regulon activity matrix' from 17,028 human pseudo-cells and 16,740 mouse pseudo-cells. The values of the matrix that corresponded to 'on' regulons in a given pseudo-cell were 1, and 0 for 'off' regulons. The regulon activity *t*-SNE maps were created using the function `tsneAUC(..., nPCs = 50, perpl = 50, aucType = "binary")` in R package SCENIC with the binary regulon activity matrix. Binary RASs for regulons were projected to the regulon activity *t*-SNEs with the function `runSCENIC_4_aucell_binarize()` in R package SCENIC. To connect regulons with cell types, we used the Wilcoxon rank-sum test to identify cell-type-specific regulons with AUC score matrices (Extended Data Fig. 11, see Source Data), and the corresponding binding motifs for regulons were obtained from JASPAR, CIS-BP or the HOCOMOCO database.

Statistics and reproducibility

In this study, we analysed around 10,000 single cells for each tissue type to ensure reproducible detection of different cell populations. Whenever donors were available, we collected 2–4 biological replicates to evaluate donor effects for different tissues. For the boxplots in Extended Data Figs. 2a, 8e, the top of the rectangle indicates the third quartile, a horizontal line near the middle of the rectangle indicates the median, and the bottom of the rectangle indicates the first quartile. A distance of 1.5 times the interquartile range (IQR) is measured out and a whisker is drawn up to the largest observed point from the data set that falls within this distance. Similarly, a distance of 1.5 times the IQR is measured out below the lower quartile and a whisker is drawn up to the lower observed point from the data set that falls within this distance. All other observed points are plotted as outliers.

Reporting summary

Further information on research design is available in the Nature Research Reporting Summary linked to this paper.

Data availability

The CNGB Nucleotide Sequence Archive accession number is CNP0000325 (<https://db.cngb.org/search/?q=CNP0000325>). The GEO accession number is GSE134355. The human DGE data are available at https://figshare.com/articles/HCL_DGE_Data/7235471. The mouse DGE data are available at https://figshare.com/articles/MCA_DGE_Data/5435866. Source Data for Figs. 2, 4 and Extended Data Figs. 2, 5, 7, 8, 9, 10, 11 are provided with the paper. HCL data can also be accessed at <http://bis.zju.edu.cn/HCL/> or <https://db.cngb.org/HCL/>.

Code availability

Detailed codes for figures are provided at <https://github.com/ggilab/HCL/>. An online R package is available for scHCL (<https://github.com/ggilab/scHCL/>).

46. Dobin, A. et al. STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* **29**, 15–21 (2013).
47. Buettner, F. et al. Computational analysis of cell-to-cell heterogeneity in single-cell RNA-sequencing data reveals hidden subpopulations of cells. *Nat. Biotechnol.* **33**, 155–160 (2015).
48. Wolf, F. A., Angerer, P. & Theis, F. J. SCANPY: large-scale single-cell gene expression data analysis. *Genome Biol.* **19**, 15 (2018).
49. Galili, T. dendextend: an R package for visualizing, adjusting and comparing trees of hierarchical clustering. *Bioinformatics* **31**, 3718–3720 (2015).
50. Suo, S. et al. Revealing the critical regulators of cell identity in the mouse cell atlas. *Cell Rep.* **25**, 1436–1445.e3 (2018).
51. Shannon, P. et al. Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Res.* **13**, 2498–2504 (2003).
52. La Manno, G. et al. RNA velocity of single cells. *Nature* **560**, 494–498 (2018).
53. Crow, M., Paul, A., Ballouz, S., Huang, Z. J. & Gillis, J. Characterizing the replicability of cell types defined by single cell RNA-sequencing data using MetaNeighbor. *Nat. Commun.* **9**, 884 (2018).

54. Guo, M., Bao, E. L., Wagner, M., Whitsett, J. A. & Xu, Y. SLICE: determining cell differentiation and lineage based on single cell entropy. *Nucleic Acids Res.* **45**, e54 (2017).
55. Grün, D. et al. De novo prediction of stem cell identity using single-cell transcriptome data. *Cell Stem Cell* **19**, 266–277 (2016).
56. Guo, J. et al. The adult human testis transcriptional cell atlas. *Cell Res.* **28**, 1141–1157 (2018).
57. Han, X. et al. Mapping human pluripotent stem cell differentiation pathways using high throughput single-cell RNA-sequencing. *Genome Biol.* **19**, 47 (2018).
58. Karamitros, D. et al. Single-cell analysis reveals the continuum of human lympho-myeloid progenitor cells. *Nat. Immunol.* **19**, 85–97 (2018).
59. Nguyen, Q. H. et al. Profiling human breast epithelial cells using single cell RNA sequencing identifies cell diversity. *Nat. Commun.* **9**, 2028 (2018).
60. Tsang, J. C. H. et al. Integrative single-cell and cell-free plasma RNA transcriptomics elucidates placental cellular dynamics. *Proc. Natl Acad. Sci. USA* **114**, E7786–E7795 (2017).
61. Villani, A. C. et al. Single-cell RNA-seq reveals new types of human blood dendritic cells, monocytes, and progenitors. *Science* **356**, eaah4573 (2017).
62. Wang, M. et al. Single-cell RNA sequencing analysis reveals sequential cell fate transition during human spermatogenesis. *Cell Stem Cell* **23**, 599–614.e4 (2018).
63. Celniker, S. E. et al. Unlocking the secrets of the genome. *Nature* **459**, 927–930 (2009).
64. Gu, Z., Gu, L., Eils, R., Schlesner, M. & Brors, B. circlize implements and enhances circular visualization in R. *Bioinformatics* **30**, 2811–2812 (2014).
65. Fuxman Bass, J. I. et al. Using networks to measure similarity between genes: association index selection. *Nat. Methods* **10**, 1169–1176 (2013).

Acknowledgements We thank G-BIO (Hangzhou), Sidansai Biotechnology (Shanghai), BGI (Shenzhen) and CNGB (Shenzhen) for supporting this project; Vazyme for supplying the customized enzymes in the study; the Core Facilities of Zhejiang University School of Medicine for technical support; the Center of Cryo-Electron Microscopy at Zhejiang University for computational support; and Y. Zhu, J. Zhu, L. Huang, L. Shao, Z. Wang, H. Huang, X. Wu, W. Lin, M. Bai, Q. Sun, X. Wu, M. Yao, F. Zhu, Z. Li, L. Huang, L. Shao, Z. Wang and X. Chen for help with sample collection. This publication is part of the Human Cell Atlas: www.humancellatlas.org/publications/. G.G. is a participant of the Human Cell Atlas Project (International), the Alliance for Atlas of Blood Cells (China), and the Cell Atlas Project (Zhejiang University Stem Cell Institute). This work was supported by the National Natural Science Foundation of China (grants 91842301, 81770188, 31722027, 31922049, 31701290, and 31871473), the National Key Research and Development Program (grants 2018YFA0107804, 2018YFA0107801, 2018YFA0800503, and 2018YFC1005003), the Zhejiang Provincial Natural Science Foundation of China (grant R17H080001), and the Fundamental Research Funds for the Central Universities (G.G.).

Author contributions The project was conceived by G.G. Tissue digestion experiments were performed by X.H., Z.Z., R.W. and Y.C. Microwell-seq experiments were performed by X.H., Z.Z., R.W., H.C., F.Y., M.J., J. Wu and S.L. Single-cell data processing, clustering and trajectory analyses were performed by L.F., H.S., J. Wang, Y.X. and C.Y. scHCL analyses and website construction were performed by H.S., Y.Z. and M.C. Cross-species and gene regulation analyses were performed by J. Wang and H.S. Immunostaining experiments were performed by H.C. Stem cell differentiation experiments were performed by H.C., X.M. and S.Z. Sequencing experiments were performed by R.L., Y.G. and M.W. Fetal tissue collections were conducted by Y.C., Y.W. and D.Z. Adult brain tissue collections were conducted by X.J., J.Z., R.Z. and H. Hu. Adult hematopoietic cell collections were conducted by H. Huang. Other adult tissue collections were conducted by H.T., W.G., T.Z., Q.Z., X.B., L.Z., C.W., T.L., J.C. and W.W.. The paper was written by G.G., X.H., Z.Z., L.F., H.S., R.W., Y.C., H.C. and J. Wang. Funding was acquired by G.G. and X.H.

Competing interests The authors declare no competing interests.

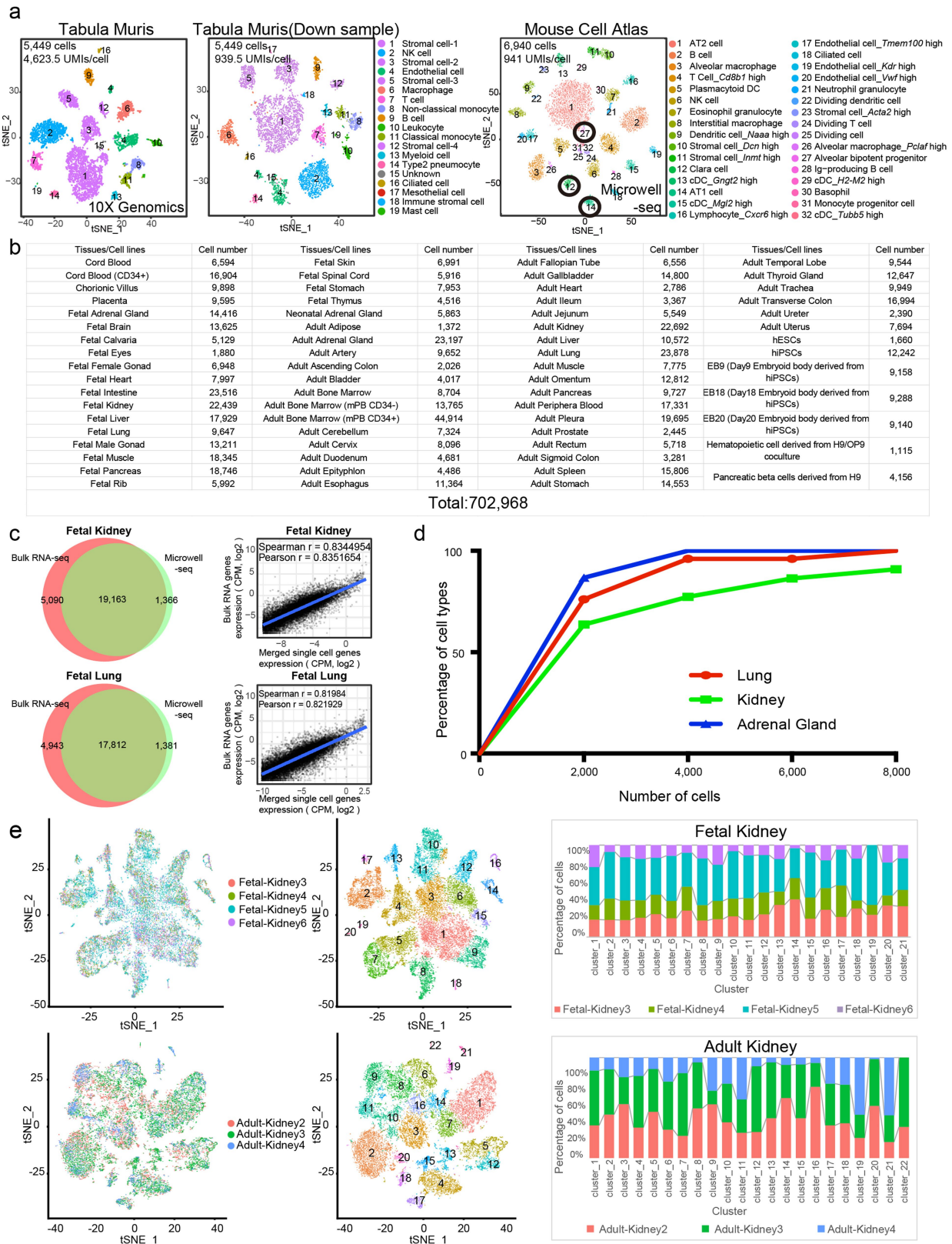
Additional information

Supplementary information is available for this paper at <https://doi.org/10.1038/s41586-020-2157-4>.

Correspondence and requests for materials should be addressed to X.H. or G.G.

Peer review information Nature thanks Berthold Gottgens, Rosario Isasi and the other, anonymous, reviewer(s) for their contribution to the peer review of this work.

Reprints and permissions information is available at <http://www.nature.com/reprints>.

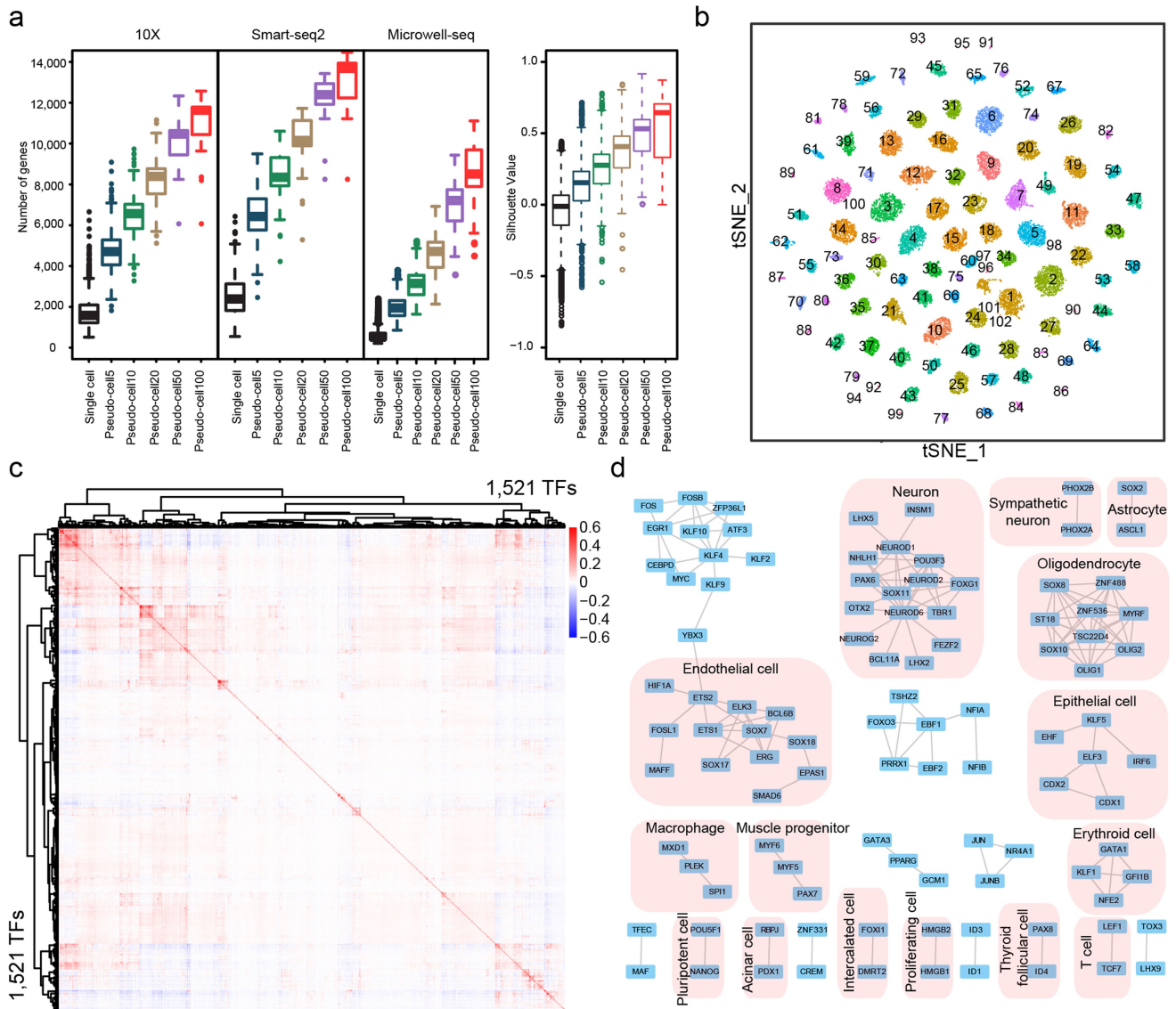


Extended Data Fig. 1 | See next page for caption.

Article

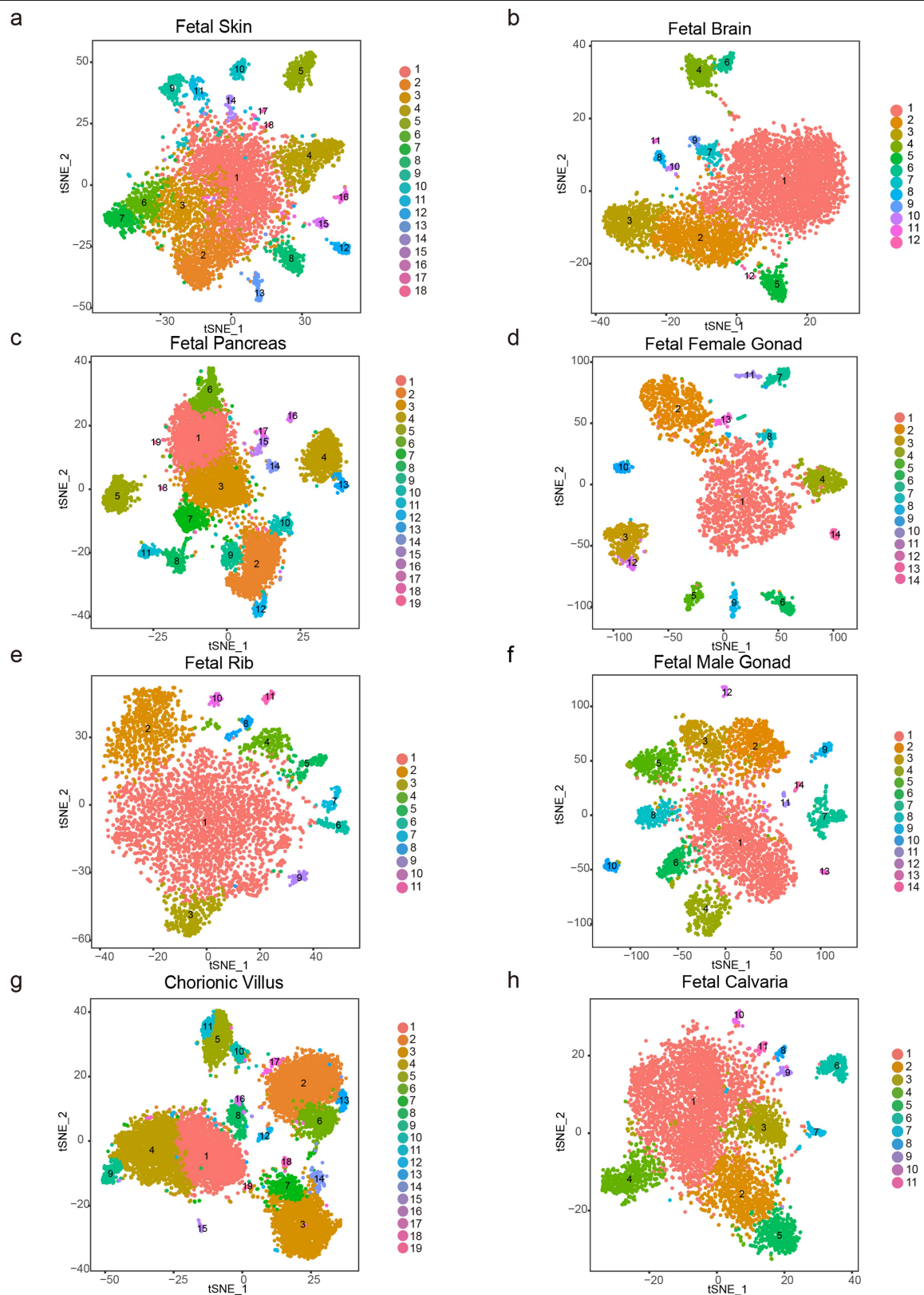
Extended Data Fig. 1 | Construction of the HCL. **a**, Comparison of *t*-SNE maps for lung data sets from Tabula Muris (10X Genomics), down-sampled Tabula Muris data (10X Genomics), and MCA data (microwell-seq). Note that downsampling in sequencing depth does not affect cell-type clusters in the Tabula Muris data. Notably, lung data from MCA (sequenced at lower depth) detects more cell-type clusters, including important lung epithelial cells such as AT1 cells, club cells and biopotent progenitors. **b**, Numbers of cells processed by 31 December 2019 at the HCL for each tissue type. **c**, Venn diagrams of gene numbers detected in bulk RNA sequencing and microwell-seq (genes with fewer than three counts are excluded). Scatter plots on the right show high correlations (more than 0.8) of average gene expression between

bulk RNA sequencing and microwell-seq. We analysed 17,058 genes for kidney and 16,910 genes for lung. **d**, The percentage of cell types recovered in sub-samples of adult lung, kidney and adrenal gland single-cell data. The major cell-type numbers in representative tissues are near plateau at around 8,000 cells; we collected more than 10,000 cells per tissue on average. **e**, A Seurat analysis of donor batch effect from four fetal kidney samples ($n = 22,439$ cells) and three adult kidney samples ($n = 22,692$ cells). The mixing of different donor single cells in each cell-type cluster suggests a relatively low batch effect in the data. The cluster contribution bar charts on the right suggest that one of the fetal kidney donors lacks C19 and one of the adult kidney donors lacks C22.



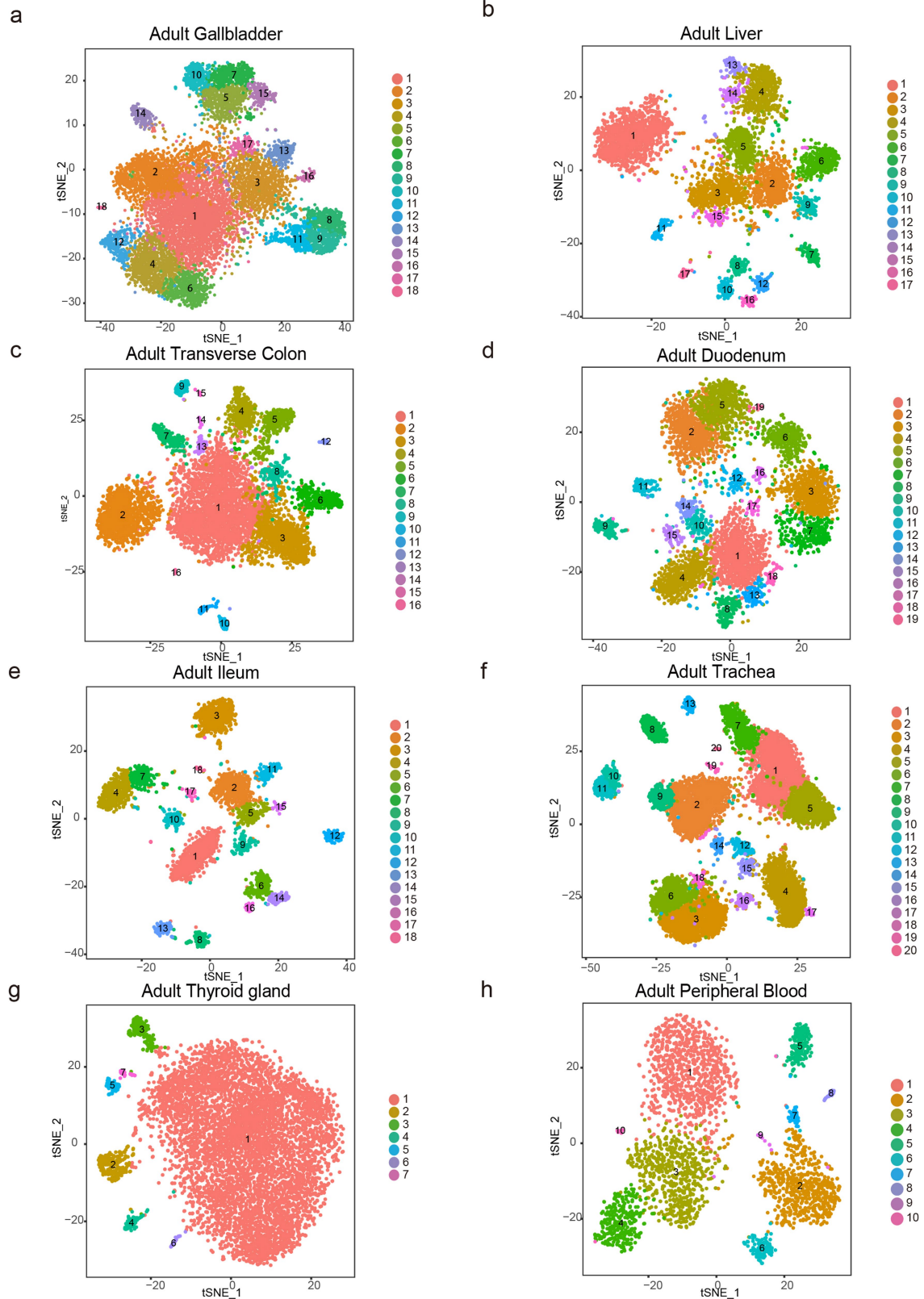
Extended Data Fig. 2 | Genetic network analysis of the HCL. a, Verification of pseudo-cell analysis for network interpretation from HCL data. Left, numbers of genes in single cell, pseudo-cell 5, pseudo-cell 10, pseudo-cell 20, pseudo-cell 50 and pseudo-cell 100 from mouse lung single-cell data^{14,15} generated by 10X Genomics, Smart-seq2 and microwell-seq. Genes were calculated in each cell or pseudo-cell. Sample sizes for each box from left to right were: 10X Genomics: 5,449, 1,089, 540, 272, 112 and 62; Smart-seq2: 1,620, 324, 158, 83, 33 and 20; microwell-seq: 6,940, 1,390, 686, 349, 142 and 81. Right, silhouette value in single cell, pseudo-cell 5, pseudo-cell 10, pseudo-cell 20, pseudo-cell 50 and pseudo-cell 100 from mouse lung single-cell data. A high

silhouette value represents good separation. Sample sizes for each box from left to right were: 6,940, 1,390, 686, 349, 142 and 81. Box plots: centre line, median; boxes, first and third quartiles of the distribution; whiskers, highest and lowest data points within $1.5 \times \text{IQR}$. **b**, t-SNE map of HCL pseudo-cell data showing improved cell-type clustering ($n = 30,053$ pseudo-cells). **c**, TF-TF correlation heatmap covering 1,521 human TFs generated using HCL pseudo-cell data. The correlation data are listed in Supplementary Table 2. **d**, Representative TF network in the HCL ($\text{PCC} > 0.5$). Note that the HCL TF network is highly related within small modules but discrete among different modules.



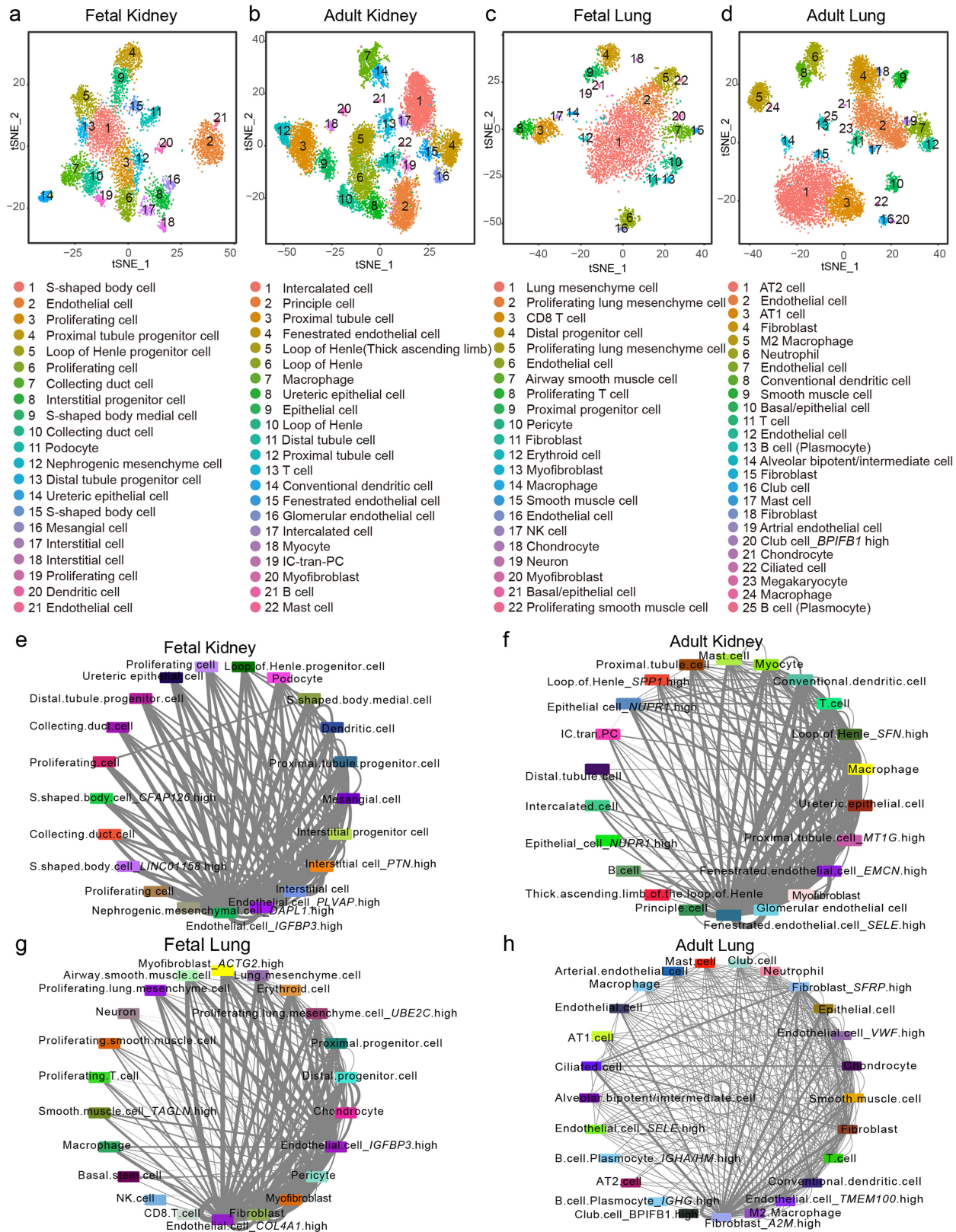
Extended Data Fig. 3 | t-SNE maps for examples of analysed fetal tissues in the HCL. t-SNE maps for single-cell data from fetal skin 2 (a; $n = 5,294$ cells), fetal brain 5 (b; $n = 5,096$), fetal pancreas 2 (c; $n = 6,939$), fetal female gonad 1

(d; $n = 2,710$), fetal rib 3 (e; $n = 4,560$), fetal male gonad 1 (f; $n = 3,358$), chorionic villus 1 (g; $n = 9,898$), and fetal calvaria 1 (h; $n = 5,129$). Cells are coloured by cell-type cluster.



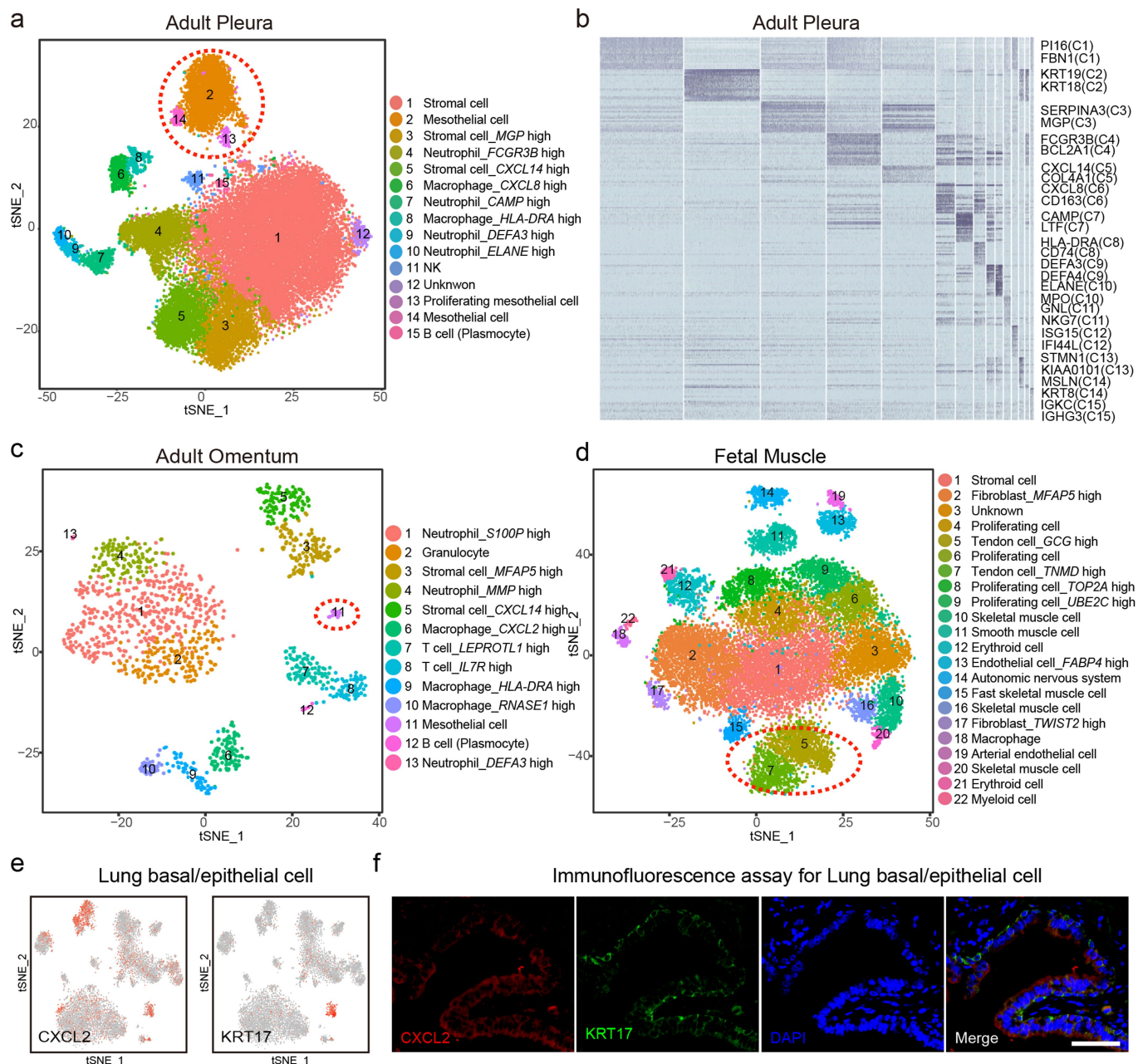
Extended Data Fig. 4 | t-SNE maps for examples of analysed adult tissues in the HCL. t-SNE maps for single-cell data from adult gallbladder 1 (**a**, $n = 9,769$ cells), adult liver 4 (**b**, $n = 4,384$), adult transverse colon 1 (**c**, $n = 5,765$), adult

duodenum 1 (**d**, $n = 4,681$), adult ileum 2 (**e**, $n = 3,367$), adult trachea 2 (**f**, $n = 9,949$), adult thyroid gland 1 (**g**, $n = 6,319$), and adult peripheral blood 1 (**h**, $n = 2,719$). Cells are coloured according to cell-type cluster.



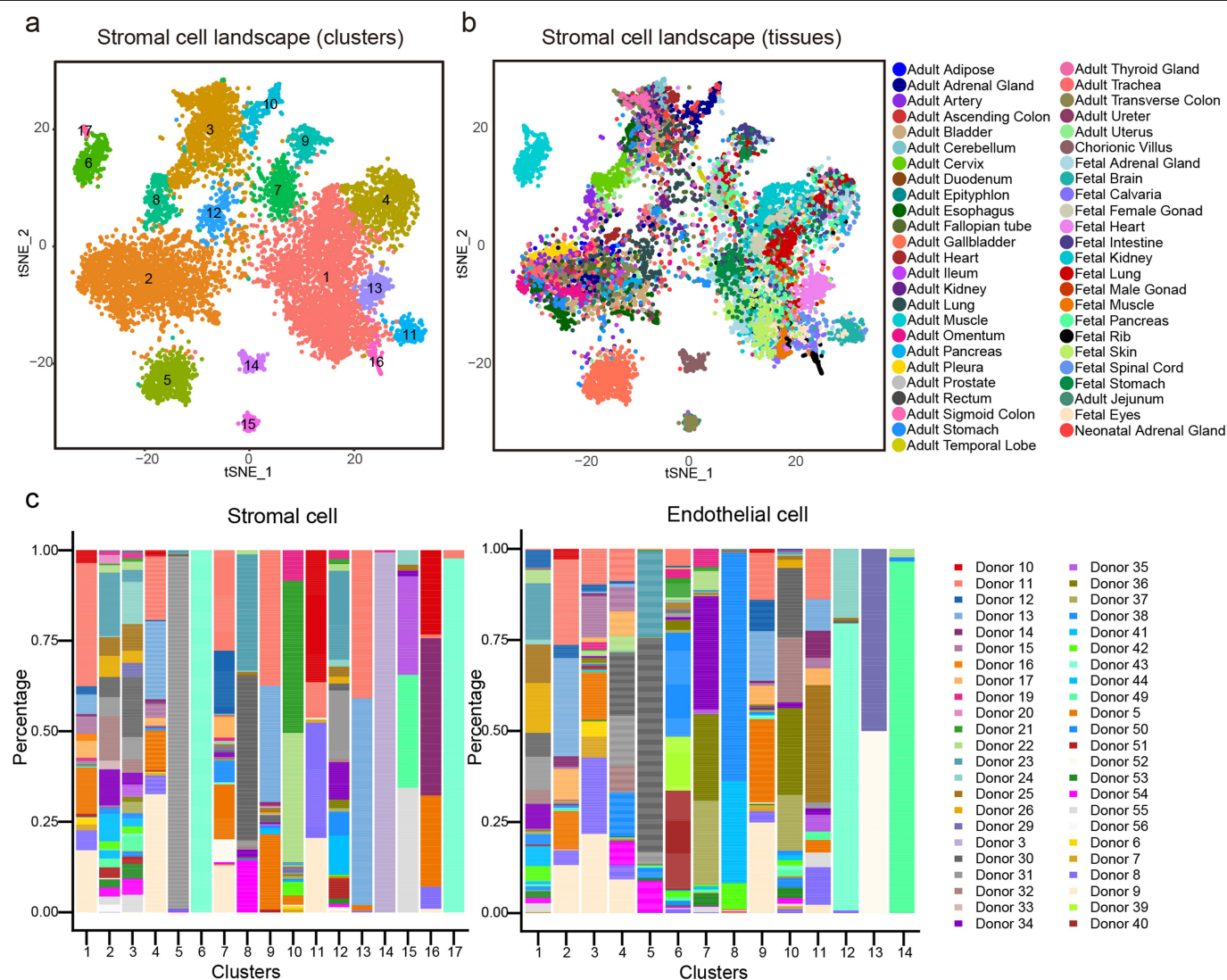
Extended Data Fig. 5 | Analysis of human lung and kidney. a, t-SNE map of fetal kidney 4 single-cell data ($n = 4,511$ cells). The experiment was replicated four times with similar results. **b**, t-SNE map of adult kidney 2 single-cell data ($n = 8,877$ cells). The experiment was replicated three times with similar results. **c**, t-SNE map of fetal lung 1 single-cell data ($n = 4,526$ cells). The experiment was replicated twice with similar results. **d**, t-SNE map of adult lung 1 single-cell data

($n = 8,426$ cells). The experiment was replicated three times with similar results. All cells in **a–d** are coloured according to cell-type cluster. **e–h**, Ligand and receptor analysis of fetal kidney 4 (**e**), adult kidney 2 (**f**), fetal lung 1 (**g**) and adult lung 1 (**h**) using the method CellPhoneDB. The colours represent cell types; line thickness indicates the degree of association between cell types.

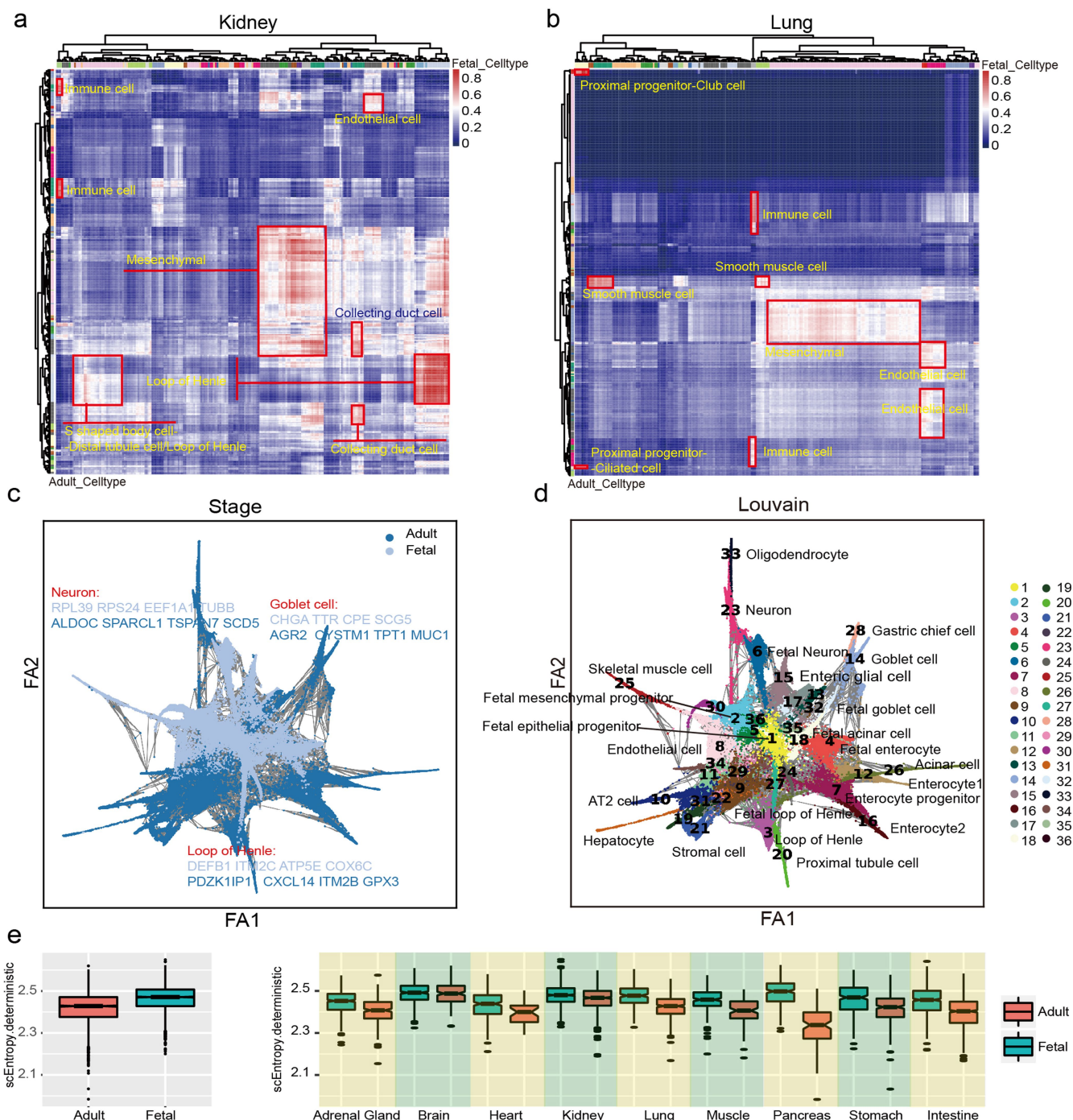


Extended Data Fig. 6 | Examples of novel populations. **a**, *t*-SNE map of adult pleura 1 single-cell data ($n = 19,695$ cells). Cells are coloured according to cell-type cluster in **a**, **c** and **d**. **b**, Gene expression heat map showing the top 20 differentially expressed genes for each cell cluster in adult pleura 1. Dark blue, high expression; light blue, low expression. Representative genes are labelled in the corresponding area on the right. **c**, *t*-SNE map of adult omentum

3 single-cell data ($n = 1,354$ cells). **d**, *t*-SNE map of fetal muscle single-cell data ($n = 18,345$ cells). **e**, Feature plot in the *t*-SNE map of adult lung 1 single-cell data ($n = 8,426$ cells). Cells are coloured according to the expression of the indicated marker genes. **f**, Immunofluorescence assay for the epithelial cell marker KRT17 and the CXC chemokine CXCL2 in human adult lung tissue. Scale bar, 50 μ m. The experiment was replicated three times with similar results.

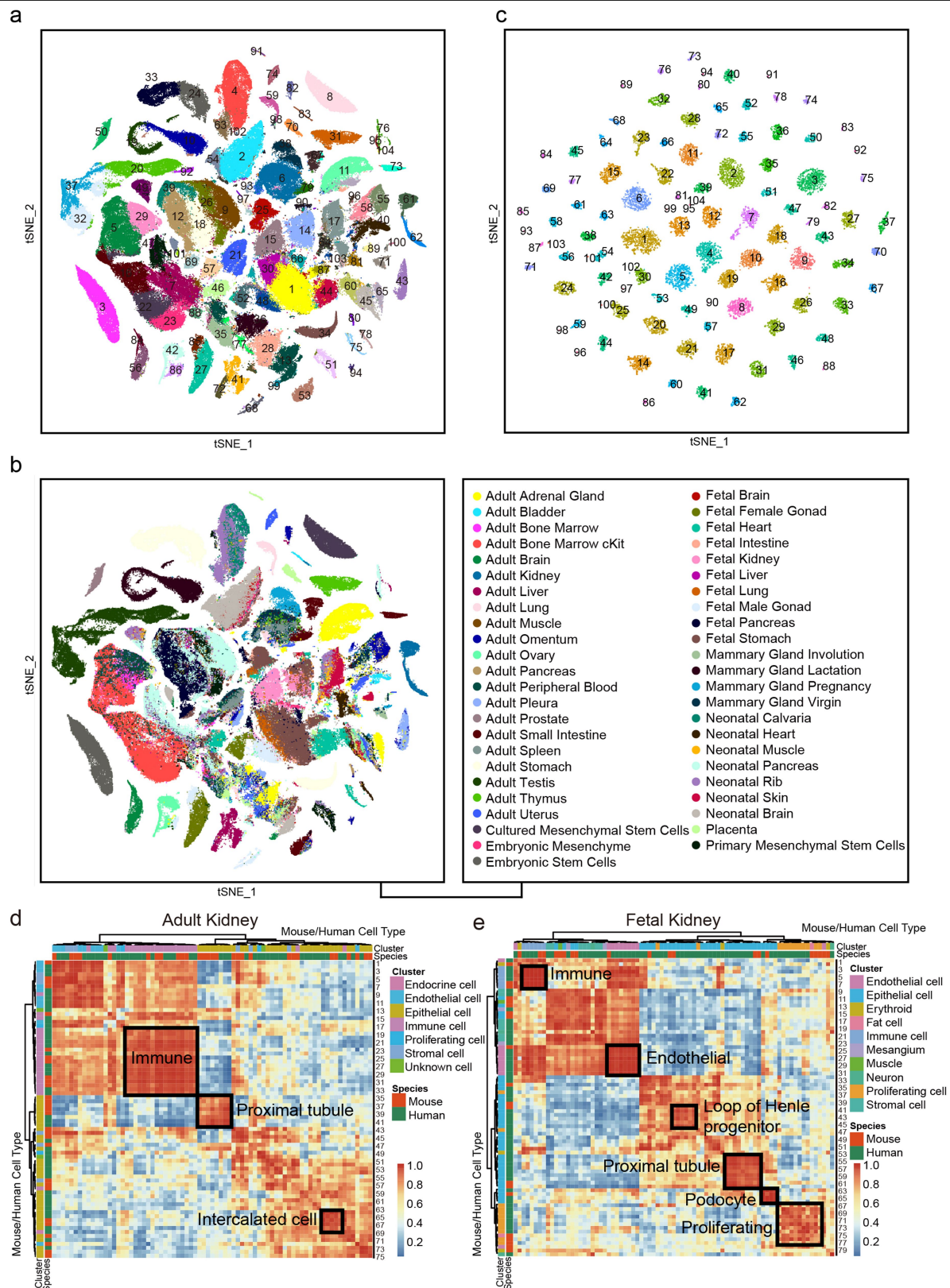


Extended Data Fig. 7 | Cross-tissue cellular network. a, b, t-SNE maps of single-cell data for human tissue-specific stromal cells ($n = 9,452$ cells). Cells are coloured according to stromal cell subtype (a) or tissue type (b). c, Bar plots showing the contributions of donors to each of the stromal and endothelial cell clusters.



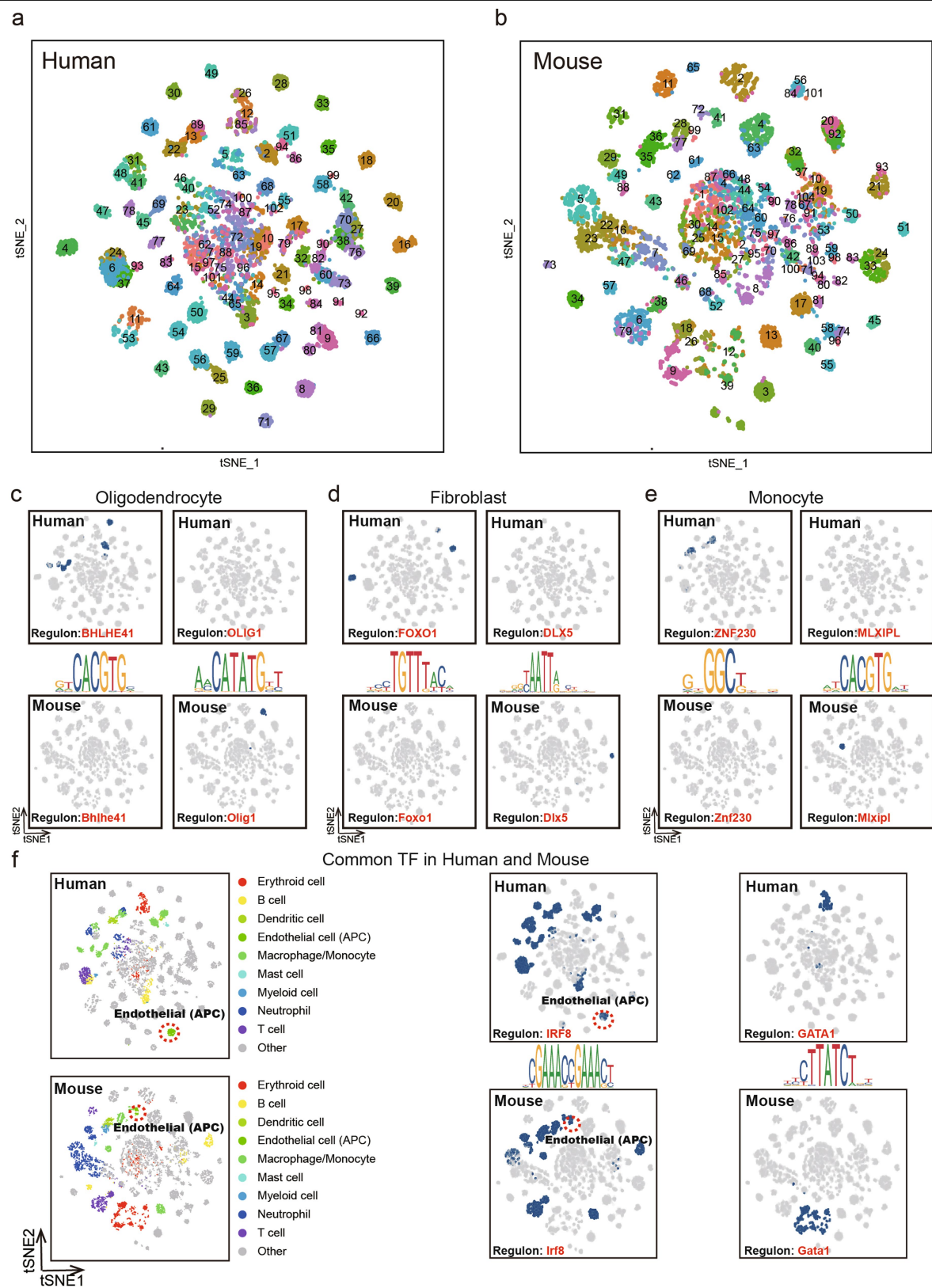
Extended Data Fig. 8 | Analysis of fetal-to-adult transition. a, Heat map showing the similarity (PCC) between cell types of adult kidney and fetal kidney. Blue, low similarity; red, high similarity. **b**, Heat map showing the similarity (PCC) between cell types of adult lung and fetal lung. Blue, low similarity; red, high similarity. **c**, **d**, Branching gene expression trajectory analysis of non-immune cells in fetal and adult human tissues using PAGA. **c**, Coloured by developmental stages; **d**, coloured by cell lineages. Differential gene expression analysis was performed for representative lineage

progression; top markers for fetal (light blue) and adult (dark blue) cells are shown. Marker lists are provided in the source data. **e**, Single-cell entropy of non-immune cells in each fetal and adult tissue pair (right) and for combined adult and fetal data (left) measured by SLICE⁵⁴. Box plots: centre line, median; boxes, first and third quartiles of the distribution; whiskers, highest and lowest data points within $1.5 \times \text{IQR}$. Sample from left to right: 2,770, 3,557, 522, 279, 340, 193, 270, 38, 885, 631, 208, 353, 187, 111, 268, 112, 294, 314, 583, 739.



Extended Data Fig. 10 | Comparison of human and mouse tissues. a, t-SNE analysis of 333,778 single cells from the MCA data, with 104 main cell-type clusters labelled. **b**, t-SNE analysis of 333,778 single cells from the MCA data, with tissue types labelled. **c**, t-SNE map of pseudo-cell 20 data for mouse ($n = 16,740$ pseudo-cells). Pseudo-cells are coloured according to cell-type cluster. **d**, Heat map showing the conserved cell types in human and mouse

adult kidney. AUROC scores were calculated from the Spearman correlation between human and mouse adult kidney pseudo-cells ($n = 1,197$ pseudo-cells). **e**, Heat map showing conserved cell types in human and mouse fetal kidney. AUROC scores were calculated from the Spearman correlation between human and mouse fetal kidney pseudo-cells ($n = 1,459$ pseudo-cells).



Extended Data Fig. 11 | See next page for caption.

Extended Data Fig. 11 | Comparison of human and mouse regulons.

a, b, Binary regulon activity *t*-SNE maps for human and mouse based on 259 human regulons (**a**) and 248 mouse regulons (**b**), created with R package SCENIC. Each dot represents a pseudo-cell of 20 in the HCL or MCA cell clusters. The *t*-SNE maps were created using binary regulon activity matrices from 17,028 human pseudo-cells and 16,740 mouse pseudo-cells. **c**, Binary RASs for human special regulon BHLHE41 and mouse special regulon Olig1 in the regulon activity *t*-SNE maps ($n = 17,028$ for human; $n = 16,740$ for mouse). **d**, Binary RASs for human special regulon FOXO1 and mouse special regulon

Dlx5 in the regulon activity *t*-SNE maps ($n = 17,028$ for human; $n = 16,740$ for mouse). **e**, Binary RASs for human special regulon ZNF230 and mouse special regulon Mxipl in the regulon activity *t*-SNE maps ($n = 17,028$ for human; $n = 16,740$ for mouse). **f**, Binary RASs for regulons IRF8/Irf8 and GATA1/Gata1 in the regulon activity *t*-SNE maps ($n = 17,028$ for human; $n = 16,740$ for mouse). Note that in the regulation of antigen-presenting endothelial cells, the IRF8/Irf8 regulon is conserved. In the regulation of erythroid cells, the GATA1/Gata1 regulon is conserved.

Reporting Summary

Nature Research wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Research policies, see [Authors & Referees](#) and the [Editorial Policy Checklist](#).

Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

n/a Confirmed

- ☐ ☒ The exact sample size (n) for each experimental group/condition, given as a discrete number and unit of measurement
- ☐ ☒ A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
- ☐ ☒ The statistical test(s) used AND whether they are one- or two-sided
Only common tests should be described solely by name; describe more complex techniques in the Methods section.
- ☒ ☐ A description of all covariates tested
- ☒ ☐ A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
- ☒ ☐ A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)
- ☒ ☐ For null hypothesis testing, the test statistic (e.g. F , t , r) with confidence intervals, effect sizes, degrees of freedom and P value noted
Give P values as exact values whenever suitable.
- ☒ ☐ For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
- ☒ ☐ For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
- ☐ ☒ Estimates of effect sizes (e.g. Cohen's d , Pearson's r), indicating how they were calculated

Our web collection on [statistics for biologists](#) contains articles on many of the points above.

Software and code

Policy information about [availability of computer code](#)

Data collection	No software was used for data collection.
Data analysis	Reads from HCL data were aligned to the Homo sapiens GRCh38 genome using STAR 2.5.2a and the DGE data matrices were obtained using the Dropseq Core Computational Protocol (available at http://mccarrolllab.org/dropseq/) with default parameters. Downstream standard procedures for filtering, variable gene selection, dimensionality reduction and clustering were performed using the Seurat 2.4.3 in R3.5.0. Scanpy 1.4.3 was used for single cell gene expression analysis, such as lineage trajectory analysis. MetaNeighbor (available at https://github.com/maggiecrow/MetaNeighbor) was used to measure the similarity of cell types. Circlize 0.4.4 was used to make circos plot. SCENIC (available at https://github.com/aertslab/SCENIC) was used to infer gene regulatory networks. Cytoscape 3.5.0 was used for network visualization. Orthofinder 2.2.6 was used to infer orthologs. Software igraph was used to perform network analysis. SLICE 0.99.0 was used for single cell entropy analysis. Velocity 0.17.13 was used to calculate RNA velocity. Cellphonedb1.0.0 was used to make ligand-receptor analysis. Detailed codes for figures are provided in github (https://github.com/ggijlab/HCL/). An online R package is available for scHCL (https://github.com/ggijlab/scHCL/).

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors/reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Research [guidelines for submitting code & software](#) for further information.

Data

Policy information about [availability of data](#)

All manuscripts must include a [data availability statement](#). This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A list of figures that have associated raw data
- A description of any restrictions on data availability

The CNGB Nucleotide Sequence Archive accession number is CNP0000325.

The GEO accession number is GSE134355.

The human DGE data are available at https://figshare.com/articles/HCL_DGE_Data/7235471.

The mouse DGE data are available at https://figshare.com/articles/MCA_DGE_Data/5435866.

Source data for Figs. 2 and 4 and Extended Data Figs. 2, 5, 7, 8, 9, 10, 11 are provided with the paper.

HCL Data can also be accessed on the website: (<http://bis.zju.edu.cn/HCL/>) and a mirror website for international users (<https://db.cngb.org/HCL/>).

Field-specific reporting

Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

☒ Life sciences

☐ Behavioural & social sciences

☐ Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see [nature.com/documents/nr-reporting-summary-flat.pdf](https://www.nature.com/documents/nr-reporting-summary-flat.pdf)

Life sciences study design

All studies must disclose on these points even when the disclosure is negative.

Sample size	702,968 single cells were analyzed in total for the first stage Human Cell landscape construction. A total of 67 human tissues and cell culture types were analyzed. In Extended Figure 1d, we estimate that the major cell-type discovery in representative tissues are near plateau at around 8000 cells. Therefore we collected more than 10000 cells per tissue on average. That makes a total of 702,968 single cells.
Data exclusions	Data points with fewer than 500 UMI were excluded. The detected transcript from a single live mammalian cell under our sequencing depth (3000 reads/cell) should be more than 500 UMI, as we have exemplified in our previous Mouse Cell Atlas paper (Han et al., Cell, 2018). Cell barcodes with less than 500 UMI usually correspond to empty beads exposed to free RNA during cell lysis, RNA capture and washing steps.
Replication	2-4 replications were done for different tissues when samples were available. The results of major cell type clusters are reproducible.
Randomization	Different single cells were randomly captured before analysis. Human samples were not randomized due to practical constraints.
Blinding	We are blinded to analyzed cell types before single cell analyses.

Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

Materials & experimental systems

n/a	Involved in the study
<input type="checkbox"/>	<input checked="" type="checkbox"/> Antibodies
<input type="checkbox"/>	<input checked="" type="checkbox"/> Eukaryotic cell lines
<input checked="" type="checkbox"/>	<input type="checkbox"/> Palaeontology
<input type="checkbox"/>	<input checked="" type="checkbox"/> Animals and other organisms
<input type="checkbox"/>	<input checked="" type="checkbox"/> Human research participants
<input checked="" type="checkbox"/>	<input type="checkbox"/> Clinical data

Methods

n/a	Involved in the study
<input checked="" type="checkbox"/>	<input type="checkbox"/> ChIP-seq
<input checked="" type="checkbox"/>	<input type="checkbox"/> Flow cytometry
<input checked="" type="checkbox"/>	<input type="checkbox"/> MRI-based neuroimaging

Antibodies

Antibodies used	Anti-HLA-DR antibody [EPR3692] (ab92511), abcam https://www.abcam.com/hla-dr-antibody-epr3692-ab92511.html CD31 Antibody Mouse Monoclonal Catalog number: 66065-1-Ig CloneNo.: 2A1E2, Proteintech https://www.ptglab.com/products/PECAM1,CD31-Antibody-66065-1-Ig.htm Cytokeratin 17 Monoclonal Antibody (E3), MA5-13539, Thermo Fisher https://www.thermofisher.com/antibody/product/Cytokeratin-17-Antibody-clone-E3-Monoclonal/MA5-13539 Anti-CXCL2 (AHP773, Bio-rad) https://www.bio-rad-antibodies.com/polyclonal/human-gro-beta-antibody-ahp773.html?f=purified The dilutions for different antibodies: [anti-HLA-DR (1:200), anti-CD31 (1:100), anti-Krt17 (1:50), and anti-CXCL2 (1:100)]
Validation	Validation are available for all antibodies from the manufacturer. Please refer to references contained in the provided links.

Eukaryotic cell lines

Policy information about [cell lines](#)

Cell line source(s)	H9 is from Wicell https://www.wicell.org/
Authentication	The H9 cell line is authenticated by multilineage differentiation experiments and mRNA-seq experiments. It shows that same self-renewal and differentiation abilities, as well as the same gene expression profiles as the other published H9 data.
Mycoplasma contamination	The cell line is negative for mycoplasma contamination.
Commonly misidentified lines (See ICLAC register)	No commonly misidentified cell lines were used

Animals and other organisms

Policy information about [studies involving animals](#); [ARRIVE guidelines](#) recommended for reporting animal research

Laboratory animals	Wild-type C57BL/6J mice were ordered from Shanghai SLAC Laboratory Animal. Mouse adult adrenal gland, omentum, pleura and stomach tissues were collected from 8-10 week-old female mice. Mouse fetal pancreas and stomach tissues were collected from E14.5 mouse embryos. All mice were housed at Zhejiang University Laboratory Animal Center in a Specific Pathogen Free (SPF) facility with individually ventilated cages. The room has controlled temperature (20-22°C), humidity (30%–70%) and light (12 hour light-dark cycle). Mice were provided ad libitum access to a regular rodent chow diet.
Wild animals	The study did not involve wild animals.
Field-collected samples	The study did not involve any samples from the field.
Ethics oversight	Mouse experiments in this study were approved by the Animal Ethics Committee of Zhejiang University; experiments conformed to the regulatory standards at Zhejiang University Laboratory Animal Center.

Note that full information on the approval of the study protocol must also be provided in the manuscript.

Human research participants

Policy information about [studies involving human research participants](#)

Population characteristics	Chinese Han population was analyzed in the study. The Han population is China's main ethnic group, with a population of about 1.3 billion, accounting for about 19% of the world's total population, and is distributed all around the world. Among 16 fetal tissue donors, 9 donors are male and 7 donors are female; all donors are between 10 weeks to 14 weeks except for 1 donor at 7 weeks and 1 donor at 26 weeks. Among 42 adult tissue donors, 22 donors are male and 20 donors are female; all donors are between 21 years to 66 years old except for one donor > 83 years and one donor < 1 year.
Recruitment	Tissue donors were recruited by collaborating doctors in the affiliated hospitals of Zhejiang University School of Medicine following local protocols. There is no potential self-selection in the volunteer tissue donation process.
Ethics oversight	The human sample collection and research conducted in this study were approved by the Research Ethics Committee of the Zhejiang University School of Medicine, Research Ethics Committee of the First Affiliated Hospital, Research Ethics Committee of the Second Affiliated Hospital and Research Ethics Committee of Women's Hospital at Zhejiang University (Approval Number: 20170029, 20180017, 20190034, 2018015, 2018507, 2018766 and 2018185). Informed consent for fetal tissue collection and research was obtained from each patient after her decision to legally terminate her pregnancy but before the abortive procedure was performed. Informed consent for collection and research of surgically removed adult tissues was obtained from each patient before the operation. Informed consent for the collection and research of tissues from deceased-organ donation was obtained from the donor family after the cardiac death of the donor. Details on donor information are provided in Supplementary Table 1. All the protocols used in this study were in strict compliance with the legal and ethical regulations of Zhejiang University School of Medicine and Affiliated Hospitals. All the protocols used in this study complied with the 'Interim Measures for the Administration of Human Genetic Resources' administered by The Ministry of Science and Technology and The Ministry of Public Health (Approval Number: 2020BAT0007).

Note that full information on the approval of the study protocol must also be provided in the manuscript.

Statin therapy is associated with lower prevalence of gut microbiota dysbiosis

<https://doi.org/10.1038/s41586-020-2269-x>

Received: 5 August 2019

Accepted: 3 April 2020

Published online: 6 May 2020

 Check for updates

Sara Vieira-Silva^{1,2,36}, Gwen Falony^{1,2,36}, Eugeni Belda^{3,4,36}, Trine Nielsen⁵, Judith Aron-Wisnewsky^{3,6}, Rima Chakaroun⁷, Sofia K. Forslund^{8,9,10,11}, Karen Assmann³, Mireia Valles-Colomer^{1,2}, Thi Thuy Duyen Nguyen^{1,2}, Sebastian Proost^{1,2}, Edi Prifti^{3,4,12}, Valentina Tremaroli¹³, Nicolas Pons¹⁴, Emmanuelle Le Chatelier¹⁴, Fabrizio Andreelli^{3,15}, Jean-Philippe Bastard^{16,17}, Luis Pedro Coelho^{9,18}, Nathalie Galleron¹⁴, Tue H. Hansen⁵, Jean-Sébastien Hulot^{19,20,21}, Christian Lewinter²², Helle K. Pedersen⁵, Benoit Quinquis¹⁴, Christine Rouault³, Hugo Roume¹⁴, Joe-Elie Salem¹⁹, Nadja B. Søndertoft⁵, Sothea Touch³, MetaCardis Consortium*, Marc-Emmanuel Dumas^{23,24}, Stanislav Dusko Ehrlich¹⁴, Pilar Galan²⁵, Jens P. Gøtze²⁶, Torben Hansen⁵, Jens J. Holst⁵, Lars Køber²², Ivica Letunic²⁷, Jens Nielsen²⁸, Jean-Michel Oppert⁶, Michael Stumvoll^{7,29}, Henrik Vestergaard⁵, Jean-Daniel Zucker^{3,4,12}, Peer Bork^{9,30,31}, Oluf Pedersen⁵, Fredrik Bäckhed^{5,13}, Karine Clément^{3,6,37} & Jeroen Raes^{1,2,37}✉

Microbiome community typing analyses have recently identified the *Bacteroides*2 (Bact2) enterotype, an intestinal microbiota configuration that is associated with systemic inflammation and has a high prevalence in loose stools in humans^{1,2}. Bact2 is characterized by a high proportion of *Bacteroides*, a low proportion of *Faecalibacterium* and low microbial cell densities^{1,2}, and its prevalence varies from 13% in a general population cohort to as high as 78% in patients with inflammatory bowel disease². Reported changes in stool consistency³ and inflammation status⁴ during the progression towards obesity and metabolic comorbidities led us to propose that these developments might similarly correlate with an increased prevalence of the potentially dysbiotic Bact2 enterotype. Here, by exploring obesity-associated microbiota alterations in the quantitative faecal metagenomes of the cross-sectional MetaCardis Body Mass Index Spectrum cohort ($n = 888$), we identify statin therapy as a key covariate of microbiome diversification. By focusing on a subcohort of participants that are not medicated with statins, we find that the prevalence of Bact2 correlates with body mass index, increasing from 3.90% in lean or overweight participants to 17.73% in obese participants. Systemic inflammation levels in Bact2-enterotyped individuals are higher than predicted on the basis of their obesity status, indicative of Bact2 as a dysbiotic microbiome constellation. We also observe that obesity-associated microbiota dysbiosis is negatively associated with statin treatment, resulting in a lower Bact2 prevalence of 5.88% in statin-medicated obese participants. This finding is validated in both the accompanying MetaCardis cardiovascular disease dataset ($n = 282$) and the independent Flemish Gut Flora Project population cohort ($n = 2,345$). The potential benefits of statins in this context will require further evaluation in a prospective clinical trial to ascertain whether the effect is reproducible in a randomized population and before considering their application as microbiota-modulating therapeutics.

Indications that alterations in the faecal microbiome are linked to the development of obesity⁵ have resulted in intense research efforts since the early days of metagenomics. However, developing a comprehensive blueprint of an obesity-associated microbiota constellation has proved challenging⁶. Although compositional observations still remain inconclusive⁷, obesity and obesity-related comorbidities have clearly

been associated with alterations in the intestinal microbiota, including lowered faecal-community richness and reduced proportional abundances of butyrate producing bacteria^{7–9}.

Cross-sectional microbiome-association studies are inherently limited regarding the inference of causality, and are potentially biased by unaccounted confounders. However, they remain highly suitable for

A list of affiliations appears at the end of the paper.

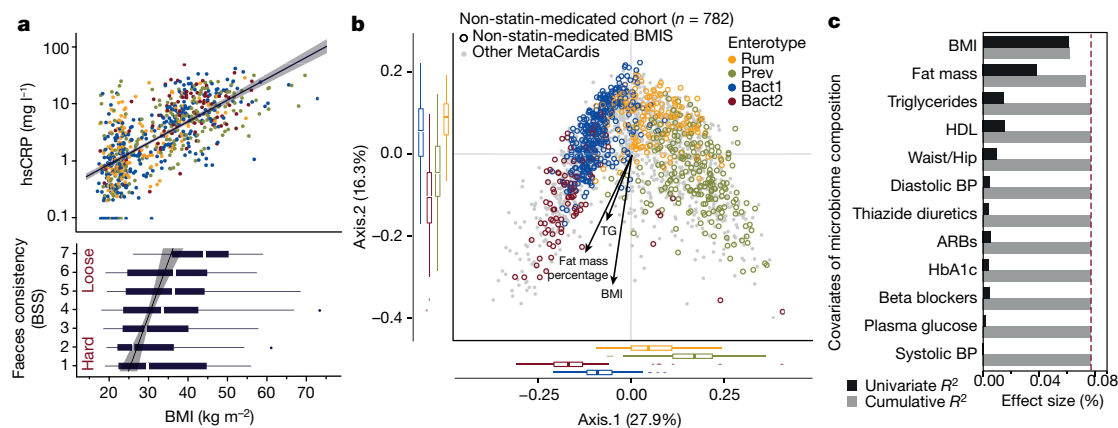


Fig. 1 | Microbiome variation in the non-statin-medicated BMIS cohort.

a, Correlations between BMI and inflammation levels (top; serum hsCRP, $n = 763$ biologically independent samples, Spearman's $\rho = 0.70$, $P_{\text{adj}} = 1.60 \times 10^{-113}$) and faeces consistency (bottom; Bristol Stool Scale (BSS), $n = 772$ biologically independent samples, Spearman's $\rho = 0.16$, $P_{\text{adj}} = 9.13 \times 10^{-6}$). Adjustment for multiple testing (P_{adj}) was performed using the Benjamini–Hochberg method. **b**, Principal coordinates analysis of inter-individual differences (genus level Bray–Curtis dissimilarity) in the microbiome profiles of the non-statin-medicated BMIS cohort (open circles, coloured by enterotype; Extended Data Fig. 4), with the rest of the MetaCardis dataset in the background ($n = 1,240$ biologically independent samples, grey dots). Arrows

represent the effect sizes of a post hoc fit of significant microbiome covariates identified in the multivariate model in **c**. **c**, Variables correlating most to microbiome compositional variation in the non-statin-medicated BMIS cohort (dbRDA, genus-level Bray–Curtis dissimilarity), either independently (univariate effect sizes in black) or in a multivariate model (cumulative effect sizes in grey). The cut-off for significant non-redundant contribution to the multivariate model is represented by the red dashed line. In **a**, **b**, the body of the box plot represents the first and third quartiles of the distribution, the line represents the median, and the whiskers extend from the quartiles to the last data point within $1.5 \times \text{IQR}$, with outliers beyond.

explorative analyses, as they enable the scale requirements imposed by the moderate effect-sizes¹⁰ to be met with relative ease. As part of the European Union MetaCardis project, a large-scale observational cohort study was set up to investigate the role of gut microorganisms in the progression of cardio-metabolic diseases through a combination of metagenomic, metabolomic and clinical approaches (<http://www.metacardis.net>). Recruitment efforts resulted in the enrolment of more than 2,000 participants (Supplementary Fig. 1) and involved, amongst others, the assembly of a transnational $n = 888$ Body Mass Index Spectrum cohort (BMIS; median BMI = 31.5 kg m^{-2} , range = 18.0–73.3; Supplementary Tables 1, 2). Faecal samples were analysed using a quantitative microbiome profiling pipeline¹ adapted for shotgun metagenomics data and were subsequently annotated with customized metabolic modules¹¹ (Supplementary Table 3). Because more than 42% of BMIS participants reported taking at least one type of medication at the time of sampling, we assessed the potential confounding effect of the most frequently disclosed therapeutics (those consumed by more than 10% of participants; Extended Data Fig. 1a, Supplementary Table 1) on the association between microbiota and obesity; this was achieved by evaluating their explanatory power on relative genus-level microbiome variation as compared with the effect-sizes of obesity parameters and host variables constituting the International Diabetes Federation consensus definition of metabolic syndrome¹² (Supplementary Table 4). Statins were identified as the drugs with largest explanatory power, contributing to genus-level microbiome variation beyond the effect of obesity-related parameters and metabolic syndrome features ($n = 869$, stepwise distance-based redundancy analysis (dbRDA), $R^2 = 0.24\%$, adjusted P value (P_{adj}) = 0.032; Extended Data Fig. 1b, c). Statin-medicated participants ($n = 106$) were most commonly prescribed simvastatin (48%; 31% atorvastatin, 21% other statins), which had an effect on microbiome variation similar to that of general statin intake (Extended Data Fig. 1d, Supplementary Table 4). To enable an—in terms of medication—least-confounded evaluation of BMI–microbiome associations, statin-medicated participants were excluded from the explorative analyses presented below.

In accordance with the premise of the analysis, within the $n = 782$ non-statin-medicated BMIS cohort (Supplementary Table 1), we found

that BMI correlated both with changes in stool consistency (higher BMI values were associated with looser stools, as assessed using the Bristol Stool Scale; $n = 772$, Spearman's $\rho = 0.16$, $P_{\text{adj}} = 9.13 \times 10^{-6}$) and with host inflammation markers (for example, serum levels of highly sensitive C-reactive protein (hsCRP), $n = 763$, Spearman's $\rho = 0.70$, $P_{\text{adj}} = 1.60 \times 10^{-113}$; Fig. 1a, Supplementary Table 5). Regarding metadata variables that define obesity or metabolic syndrome, only BMI, fat mass percentage and serum fasting triglycerides were found to explain a both significant and non-redundant fraction of compositional microbiome variation ($n = 764$, stepwise dbRDA, $R^2 = 6.22\%$ ($P_{\text{adj}} = 1.0 \times 10^{-4}$), 1.15% (1.0×10^{-4}) and 0.39% (0.009), respectively; Fig. 1b, c, Supplementary Table 6). All three covariates correlated with microbiome gene richness ($n = 771$, Spearman's $\rho = -0.45$ to -0.26 , $P_{\text{adj}} = 4.0 \times 10^{-39}$ to 1.6×10^{-13}), a proxy for microbial biodiversity proposed as a marker of metabolic health in obese individuals⁸, and with faecal microbial load ($n = 771$, Spearman's $\rho = -0.17$ to -0.13 , $P_{\text{adj}} = 4.1 \times 10^{-6}$ to 3.1×10^{-4} ; Extended Data Fig. 2, Supplementary Table 7, Supplementary Fig. 2). Additionally, BMI, fat mass percentage and triglycerides could all be linked to quantitative variation in specific microbiome features, in terms of composition as well as metabolic potential (Supplementary Table 8). Notable associations included the decrease in *Akkermansia*¹³—which is associated with metabolic health—with increasing BMI ($n = 432$, Spearman's $\rho = -0.23$, $P_{\text{adj}} = 6.8 \times 10^{-9}$), alongside an increase in, for example, *Acidaminococcus* spp. ($n = 163$, Spearman's $\rho = 0.23$, $P_{\text{adj}} = 5.8 \times 10^{-9}$), a genus that has previously been linked to body mass in a large Korean cohort¹⁴. The abundance of *Faecalibacterium*—a genus with potential anti-inflammatory properties¹⁵—was negatively correlated with all three parameters assessed, but was most closely associated with serum triglyceride levels ($n = 753$, Spearman's $\rho = -0.16$, $P_{\text{adj}} = 2.5 \times 10^{-4}$). Covariation patterns between BMI, fat mass percentage or triglyceride levels and gut-microbial metabolic modules consisted nearly exclusively of negative correlations (Supplementary Table 8), reflecting the accompanying overall decrease in total microbial load (Supplementary Table 7). Among the features that decrease with all three variables, we highlight that the variation in the butyryl-CoA-acetate CoA-transferase pathway¹⁶—the most common butyrate production pathway in colon bacteria ($n = 771$, Spearman's $\rho = -0.27$ to

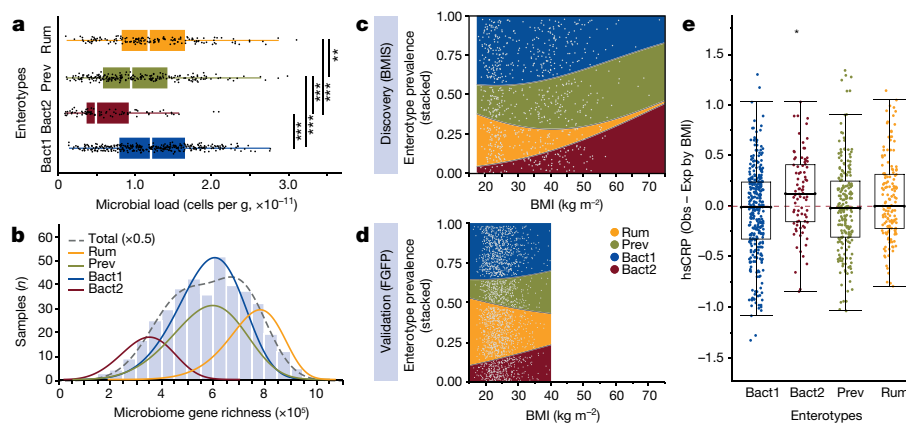


Fig. 2 | Characterization of enterotypes and variation in prevalence with BMI in the non-statin-medicated BMIS cohort. **a**, Distribution of faecal microbial loads across enterotypes, showing decreased microbial density in the Bact2 enterotype ($n = 771$ biologically independent samples, Kruskal–Wallis with post hoc Dunn test, $***P_{\text{adj}} < 0.001$; $**P_{\text{adj}} < 0.01$; Supplementary Table 9). **b**, Distribution of gene richness between enterotypes, with low richness samples corresponding to the Bact2 community constellation ($n = 782$ biologically independent samples). **c**, Variation in the prevalence of enterotypes with the BMI of individuals, showing the significant increase in Bact2 prevalence with obesity ($n = 782$ biologically independent samples, binomial logistic regression, Bact2 relative risk = 1.05, $P = 1.2 \times 10^{-7}$). Coloured areas represent the stacked enterotype prevalence along the BMI gradient, with lines provided by multivariate logistic regression of enterotypes by BMI,

and data points (light grey) jittered at the corresponding BMI. **d**, Validation of the association between BMI and Bact2 prevalence in the independent FGFP dataset ($n = 2,051$ participants, excluding statin-medicated individuals; binomial logistic regression, relative risk = 1.03, $P = 9.4 \times 10^{-3}$). **e**, Inflammatory levels are higher in Bact2 carriers than would be expected on the basis of BMI, as shown by the distribution of residuals of the linear regression between serum CRP and BMI ($n = 763$ biologically independent samples, one-sample location test (dotted line, null hypothesis; mean = 0), Cohen's $d = 0.27$, $*P_{\text{adj}} = 0.018$). In **a**, **e**, the body of the box plot represents the first and third quartiles of the distribution, the line represents the median, and the whiskers extend from the quartiles to the last data point within $1.5 \times \text{IQR}$, with outliers beyond.

-0.20 , $P_{\text{adj}} = 3.1 \times 10^{-13}$ to 6.0×10^{-8} ; Extended Data Fig. 3a–c)—is in line with previous reports linking this pathway with metabolic health⁸. The metabolism of microbiota-derived butyrate by colonocytes is essential for the maintenance of hypoxic conditions within the colon environment¹⁷, and the disruption of microbial butyrate production has been suggested to induce low-diversity gut microbiota dysbiosis¹⁸.

To investigate a potential association between BMI and the prevalence of faecal microbiome community constellations, we enterotyped the BMIS cohort using Dirichlet multinomial mixtures on genus-level molecular operational taxonomic unit (mOTU) profiles. By applying probabilistic models to group samples that potentially originate from the same community, stratification based on Dirichlet multinomial mixtures reproducibly identifies microbiome constellations across datasets without making any claims regarding the putative discrete nature of the strata detected. Our analyses confirmed previous reports of microbiome variation centred around four enterotypes^{12,19} (Fig. 1b, Extended Data Fig. 4a, b), hereafter termed Ruminococcaceae (Rum), *Bacteroides*1 (Bact1), *Bacteroides*2 (Bact2) and *Prevotella* (Prev) on the basis of their respective genus-level proportional abundance profiles (Extended Data Fig. 4c). Cell counts differed between enterotypes¹, with the low-richness Bact2 samples ($n = 782$, Kruskal–Wallis, $\chi^2 = 325.65$, $P_{\text{adj}} = 5.5 \times 10^{-70}$) also exhibiting the lowest microbial loads ($n = 771$, Kruskal–Wallis, $\chi^2 = 80.14$, $P_{\text{adj}} = 2.9 \times 10^{-17}$; Fig. 2a, Supplementary Table 9).

A quantitative compositional and functional analysis of the differences between enterotypes aligned with previous reports¹¹ (Supplementary Table 10). Further to the findings highlighted above, we found that Bact2 communities displayed the lowest abundances of *Akkermansia* ($n = 771$, Kruskal–Wallis, $\chi^2 = 141.12$, $P_{\text{adj}} = 2.0 \times 10^{-29}$) and of *Faecalibacterium* ($n = 771$, Kruskal–Wallis, $\chi^2 = 112.73$, $P_{\text{adj}} = 1.7 \times 10^{-23}$), as well as a decreased butyrate production potential ($n = 771$, Kruskal–Wallis, $\chi^2 = 167.12$, $P_{\text{adj}} = 4.7 \times 10^{-35}$; Extended Data Fig. 3d). Whereas no significant differences in *Acidaminococcus* levels could be noted between enterotypes ($n = 771$, Kruskal–Wallis, $\chi^2 = 6.47$, $P_{\text{adj}} = 0.12$), taxa such as *Eggerthella*—a genus that is considered part of a normal microbiota but is also linked to gastrointestinal infections as well as bacteraemia²⁰—was

found to occur in higher absolute numbers against the background of overall reduced microbial load, as observed in Bact2 communities ($n = 771$, Kruskal–Wallis, $\chi^2 = 224.95$, $P_{\text{adj}} = 4.1 \times 10^{-47}$; Extended Data Fig. 5a, b, Supplementary Table 10). Co-abundance gene group analyses additionally indicated enterotype differentiation at the species level (Supplementary Table 11). For example, in Bact2-enterotyped communities, the *Bacteroides* fraction was observed to be proportionally depleted in *Bacteroides caccae* ($n = 768$, Kruskal–Wallis, $\chi^2 = 78.40$, $P_{\text{adj}} = 1.3 \times 10^{-15}$) and *Bacteroides cellulosilyticus* ($n = 768$, Kruskal–Wallis, $\chi^2 = 64.79$, $P_{\text{adj}} = 5.3 \times 10^{-13}$) when compared with Rum, Prev and Bact1 samples. By contrast, it seemed to be enriched in *Bacteroides fragilis* ($n = 768$, Kruskal–Wallis, $\chi^2 = 65.26$, $P_{\text{adj}} = 3.5 \times 10^{-11}$; Extended Data Fig. 6, Supplementary Table 11), which is considered to be among the most pathogenic and immunomodulatory of the *Bacteroides* species^{21,22}.

The prevalence of enterotypes along a gene-richness axis in the non-statin-medicated cohort confirmed previous observations of a bimodal distribution⁸; however, the Bact2 community type enabled further refinement of richness stratifications through the deconvolution of overlapping peaks (Fig. 2b). The prevalence of Bact2 was found to increase with BMI, from 3.90% among lean or overweight participants ($\text{BMI} < 30$) to 17.73% among obese participants ($\text{BMI} \geq 30$) ($n = 782$, binomial logistic regression, relative risk = 1.05, $P = 1.2 \times 10^{-7}$, where relative risk can be interpreted as the scale factor necessary to obtain the prevalence of the Bact2 enterotype after a unit increase in BMI; Fig. 2c; Supplementary Table 12). Notwithstanding methodological differences, this finding was validated in the independent, amplicon-sequenced Flemish Gut Flora Project¹⁰ dataset (FGFP, $n = 2,051$; excluding statin-medicated participants; binomial logistic regression, relative risk = 1.03, $P = 9.3 \times 10^{-3}$; Fig. 2d). In line with previous findings from the FGFP², Bact2 hosts from the BMIS cohort displayed more pronounced systemic inflammation levels when compared to non-Bact2 participants, here assessed through serum hsCRP concentrations (Kruskal–Wallis, $\chi^2 = 48.61$, $P = 1.37 \times 10^{-10}$; Extended Data Fig. 7a; Supplementary Table 13). Notably, the inflammatory tone of Bact2 hosts exceeded the levels anticipated on the basis of their obesity status ($n = 86$, one-sample location test on residuals of non-statin-medicated BMIS linear regression

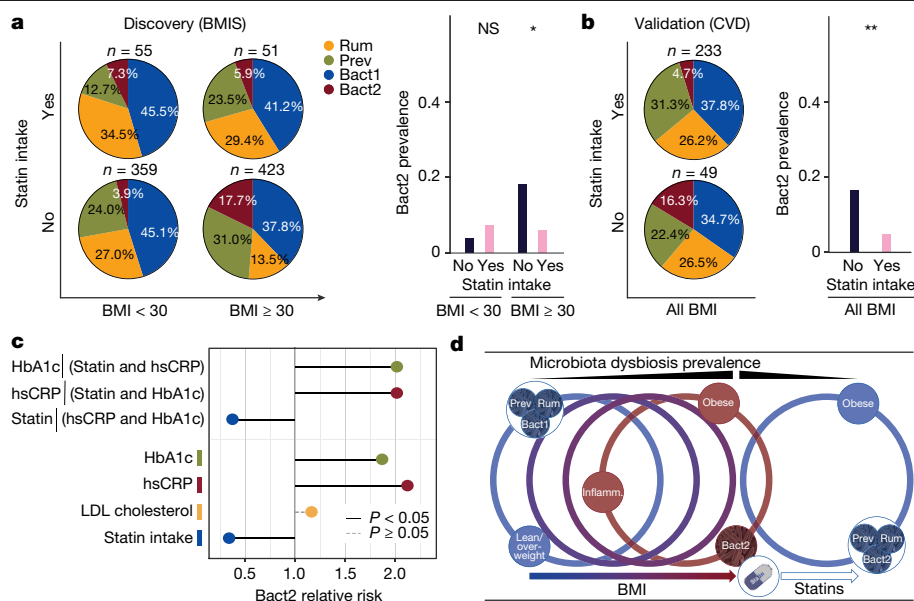


Fig. 3 | Association between the prevalence of the Bact2 enterotype, obesity and statin intake. a, Bact2 prevalence in obese (BMI ≥ 30) compared with lean and overweight (BMI < 30) individuals in the BMIS cohort (n = 888), stratified according to statin medication status. Statin-mediated obese individuals display a significantly lower prevalence of Bact2 as compared with non-statin-medicated individuals (bar plots; statin-medicated versus non-statin-medicated, 5.88% versus 17.73%, n = 888 biologically independent samples, Fisher's two-tail exact test, log likelihood = -2.88, *P = 0.028). **b**, The lower Bact2 prevalence among statin-medicated compared with non-statin-medicated individuals is validated in the MetaCardis cardiovascular disease cohort, comprising n = 282 non-diabetic patients with cardiovascular disease (CVD; statin-medicated versus non-statin-medicated, 4.72% versus 16.33%, n = 282 biologically independent samples, Fisher's two-tail exact test, log likelihood = -3.47, **P = 0.008). **c**, Relative risk of obese BMIS individuals (n = 474 participants) harbouring a Bact2 enterotype as a function of statin intake and serum biomarkers for potential (side) effects of statins (lipidemic control (LDL-cholesterol), inflammatory modulation (hsCRP) and glucose

regulation (HbA1c). Variables were modelled independently or together in univariate or multivariate models, respectively (Supplementary Table 19). The latter suggests that statin intake remains associated with a reduction in dysbiosis risk after partialing-out hsCRP and HbA1c (n = 462 biologically independent samples, multivariate binomial logistic regression, Statin | (hsCRP and HbA1c) relative risk = 0.36, $P_{\text{adj}} = 0.039$). Adjustment for multiple testing (P_{adj}) was performed on univariate tests using the Benjamini-Hochberg method (represented by black lines when significant ($P_{\text{adj}} < 0.05$), or otherwise a dashed grey line ($P_{\text{adj}} = 0.15$)). **d**, Graphical summary of the main results regarding the prevalence of the Bact2 enterotype, BMI and statin intake. In the present BMIS cohort, we identify Bact2 as an inflammation-associated microbiome community constellation, with increasing prevalence along a BMI gradient in non-statin-medicated individuals. Statin therapy is associated with attenuated inflammation and a Bact2 prevalence comparable to that observed among lean and overweight subjects. Circles represent individual host configurations in terms of body mass, microbiota community type, and inflammation status.

between hsCRP and BMI, Cohen's $d = 0.27$, $P_{\text{adj}} = 0.018$; Fig. 2e, Extended Data Fig. 7b, Supplementary Table 14). In a multivariate model, the BMI and the Bact2 carrier status of the participants both provided a non-redundant contribution to increased systemic inflammation levels, corresponding to a 1.04 (n = 763, linear multivariate model, $P_{\text{adj}} = 2.2 \times 10^{-16}$) and a 1.16 ($P_{\text{adj}} = 0.004$) unit increase risk in serum hsCRP levels, respectively (Supplementary Table 15). These observations support the qualification of the Bact2 microbiota configuration as an (low-grade) inflammation-associated, potentially dysbiotic enterotype^{1,2}.

Whether initiating or sustaining pro-inflammatory processes and metabolic derailment, countering dysbiosis of the gut ecosystem has been suggested to contribute to the maintenance of host health and the containment of obesity-related comorbidities. However, no effective microbiome modulation strategy has yet been established. Here, within the limitations of the cross-sectional study design, we identify statin treatment as a potential lever in the management of dysbiosis. In contrast to the findings from the non-statin-medicated participants, we observed that Bact2 prevalence no longer significantly increased with BMI in statin-medicated individuals (n = 106, binomial logistic regression, relative risk = 1.03, $P = 0.60$). Among obese individuals, only 5.88% of statin-medicated individuals were enterotyped as Bact2, compared with 17.73% of non-statin-medicated participants (Fisher's two-tail exact test, log likelihood = -2.88, $P = 0.028$; Fig. 3a, Supplementary Table 16). When exploring whether accounted clinical parameters, anticipated treatment responses, co-medication or key

microbiome covariates¹⁰ could be associated with the observed differences in Bact2 prevalence, we noted that statin-medicated obese participants displayed ameliorated lipid metabolism (low-density lipoprotein (LDL)-cholesterol, n = 473, Mann-Whitney U-test, $r = -0.17$, $P_{\text{adj}} = 0.002$) and inflammation status (hsCRP, n = 462, Mann-Whitney U-test, $r = -0.23$, $P_{\text{adj}} = 8.4 \times 10^{-6}$; Supplementary Table 17)—both expected outcomes of statin therapy²³. Besides minor differences in the incidence of concomitant drug intake (notably aspirin intake being more prevalent among statin-medicated participants; n = 474, Fisher's two-tailed exact test, log likelihood = -17.36, $P_{\text{adj}} = 2.2 \times 10^{-7}$) and glucose metabolism (lower HbA1c levels among non-statin-medicated participants, n = 474, Mann-Whitney U-test, $r = 0.17$, $P_{\text{adj}} = 0.001$)—the latter being a known side effect of statin treatment²⁴—the statin-medicated subcohort was characterized as older (median age statin-medicated versus non-statin-medicated, 61 versus 47; n = 474, Mann-Whitney U-test, $r = 0.34$, $P_{\text{adj}} = 1.4 \times 10^{-11}$) and less obese (BMI 33.5 versus 40.8; n = 474, Mann-Whitney U-test, $r = -0.25$, $P_{\text{adj}} = 2.1 \times 10^{-6}$). However, among these significant covariates, and excluding variables that reflect pleiotropic effects of statins—that is, levels of LDL-cholesterol and inflammation markers—only statin intake and blood HbA1c levels were shown to have a significant, non-redundant explanatory power for Bact2 prevalence (Supplementary Table 18), with the latter being associated with an increased probability of Bact2 carrier status (n = 472, multivariate binomial logistic regression, statin intake relative risk = 0.31, $P_{\text{adj}} = 0.013$; HbA1c relative risk = 2.00, $P_{\text{adj}} = 0.009$). Although 41% of

BMIS participants reported taking non-statin drugs, (co-)medication status did not affect the outcome of Bact2 prevalence analyses in obese participants (Extended Data Fig. 8). Low prevalence of the Bact2 enterotype among statin-medicated individuals was validated in the accompanying MetaCardis cardiovascular disease dataset (non-diabetic patients with cardiovascular disease (CVD); Bact2 prevalence among statin-medicated versus non-statin-medicated participants, 4.72% versus 16.33%; $n = 282$, Fisher's two-tailed exact test, log likelihood = -3.47 , $P = 0.008$; Fig. 3b, Supplementary Table 16). Here—and in accordance with observations in non-CVD disease cohorts^{1,2}—increased Bact2 prevalence was not limited to obese non-statin-medicated patients with CVD, but could also be noted within the non-statin-medicated lean and overweight subgroup. In the independent FGFP dataset—which targets an average representation of a Western population, and therefore covers a narrower BMI spectrum ($n = 2,345$; median BMI = 24.2, range = 16–40)—we confirmed lowered Bact2 prevalence among statin-medicated individuals given their BMI ($n = 2,345$, multivariate binomial logistic regression, Statin | BMI, relative risk = 0.72, $P_{\text{adj}} = 0.045$; Extended Data Fig. 9, Supplementary Table 16). Additional evidence—which is indicative of causality in statin-associated microbiota variation—is provided by a recent small-scale intervention study in a rat model, which demonstrates reversion of microbiota alterations induced by a high-fat diet and hypercholesterolemia upon treatment with atorvastatin, resulting in an increased microbiome richness²⁵. Although caution should be applied when extrapolating findings from the rodent microbiome to a human context, these results do demonstrate directionality in statin–microbiota associations, although the effect of atorvastatin (31% of statin-medicated participants) in the present BMIS cohort did not reach statistical significance (Extended Data Fig. 1, Supplementary Table 4).

The cross-sectional nature of the MetaCardis dataset did not enable us to establish a causal chain of events that lead to a lower prevalence of the Bact2 enterotype among statin-medicated individuals. Given the putatively independent effects of statin therapy on levels of serum hsCRP and LDL-cholesterol²³, we modelled the association of both variables with Bact2 prevalence for obese participants in the BMIS cohort. Although no significant effect of LDL-cholesterol concentrations was found ($n = 473$, univariate binomial logistic regression, LDL-cholesterol relative risk = 1.16, $P_{\text{adj}} = 0.15$), lower hsCRP levels were associated with a lower prevalence of the Bact2 enterotype ($n = 462$, univariate binomial logistic regression, hsCRP relative risk = 2.11, $P_{\text{adj}} = 0.003$; Supplementary Table 19). A multivariate model for the prediction of Bact2 prevalence—which covers treatment (statin intake), treatment outcome (hsCRP levels), as well as side effects of treatment (HbA1c concentrations)—indicated a significant additive contribution of statin therapy to the reduction of dysbiosis risk ($n = 462$, multivariate binomial logistic regression, Statin | (hsCRP and HbA1c) relative risk = 0.36, $P_{\text{adj}} = 0.039$; Fig. 3c, Extended Data Fig. 10, Supplementary Table 19); this suggests that the effect of statins is greater than solely the attenuating effect on the inflammation status of the host. Nevertheless, the pleiotropic effect of statins on microbiome community constellations seemed to be closely associated with a concomitant effect on host inflammation levels. At this point, at least two mechanistic interpretations of our observations—or a combination of both—remain possible (Fig. 3d). On one hand, aligning with the microbiota–inflammation hypothesis, statins could counteract the microbial contribution to inflammatory and metabolic obesity comorbidities through (in)direct modulation of the microbiota. Consistent with this, *in vitro* studies have demonstrated that statins affect the growth of several gut microorganisms²⁶. Conversely, the demonstrated anti-inflammatory effects of statins could alleviate gut host–microbial interactions and enable the subsequent development of enterotypes that are not associated with inflammation. However, it should be stressed that the cross-sectional design of our study does not allow us to rule out potential confounding by indication (lower Bact2 prevalence resulting from the specific condition that

prompted statin prescription) or by unaccounted diagnosis-associated diet or lifestyle alterations (participants adopting health-promoting and/or microbiota-modulating activities complementary to statin therapy).

For many years, strategies for the modulation of microbiota have revolved around (next-generation) probiotics and prebiotics—introducing or promoting the growth of beneficial bacteria or bacterial consortia. It is only recently that a revived interest in the effect of small molecules and drugs on the colon ecosystem, as well as individual faecal isolates, has been noted^{26,27}. Although we cannot rule out a potential effect of unaccounted confounders, nor can we infer causality from the associations observed, our analyses reveal that statin therapy is linked with a lowered prevalence of a pro-inflammatory microbial community type in obese individuals. Our results align well with previous, sparse reports of a beneficial effect of statins in pathologies in which a role of the gut microbiota has been postulated²⁸—including interventional²⁹ as well as epidemiological³⁰ evidence in Crohn's disease, a condition that has previously been linked to a high prevalence of Bact2^{1,2}. Within the limitations of the cross-sectional nature of the cohorts analysed—and emphasizing the need for interventional follow-up research using a randomized, double-blind, placebo-control study design to exclude potential confounding by indication—our findings suggest statins as a possible target for the development of future drug-based strategies for the modulation of the intestinal microbiota.

Online content

Any methods, additional references, Nature Research reporting summaries, source data, extended data, supplementary information, acknowledgements, peer review information; details of author contributions and competing interests; and statements of data and code availability are available at <https://doi.org/10.1038/s41586-020-2269-x>.

- Vandeputte, D. et al. Quantitative microbiome profiling links gut community variation to microbial load. *Nature* **551**, 507–511 (2017).
- Vieira-Silva, S. et al. Quantitative microbiome profiling disentangles inflammation- and bile duct obstruction-associated microbiota alterations across PSC/IBD diagnoses. *Nat. Microbiol.* **4**, 1826–1831 (2019).
- Probert, C. S., Emmett, P. M. & Heaton, K. W. Some determinants of whole-gut transit time: a population-based study. *QJM* **88**, 311–315 (1995).
- Ford, E. S. Body mass index, diabetes, and C-reactive protein among U.S. adults. *Diabetes Care* **22**, 1971–1977 (1999).
- Turnbaugh, P. J. et al. An obesity-associated gut microbiome with increased capacity for energy harvest. *Nature* **444**, 1027–1031 (2006).
- Sze, M. A. & Schloss, P. D. Looking for a signal in the noise: revisiting obesity and the microbiome. *MBio* **7**, e01018-16 (2016).
- Walters, W. A., Xu, Z. & Knight, R. Meta-analyses of human gut microbes associated with obesity and IBD. *FEBS Lett.* **588**, 4223–4233 (2014).
- Le Chatelier, E. et al. Richness of human gut microbiome correlates with metabolic markers. *Nature* **500**, 541–546 (2013).
- Karlsson, F. H. et al. Symptomatic atherosclerosis is associated with an altered gut metagenome. *Nat. Commun.* **3**, 1245 (2012).
- Falony, G. et al. Population-level analysis of gut microbiome variation. *Science* **352**, 560–564 (2016).
- Vieira-Silva, S. et al. Species-function relationships shape ecological properties of the human gut microbiome. *Nat. Microbiol.* **1**, 16088 (2016).
- Alberti, K. G. M. M., Zimmet, P. & Shaw, J. Metabolic syndrome—a new world-wide definition. A consensus statement from the International Diabetes Federation. *Diabet. Med.* **23**, 469–480 (2006).
- Depommier, C. et al. Supplementation with *Akkermansia muciniphila* in overweight and obese human volunteers: a proof-of-concept exploratory study. *Nat. Med.* **25**, 1096–1103 (2019).
- Yun, Y. et al. Comparative analysis of gut microbiota associated with body mass index in a large Korean cohort. *BMC Microbiol.* **17**, 151 (2017).
- Quévrain, E. et al. Identification of an anti-inflammatory protein from *Faecalibacterium prausnitzii*, a commensal bacterium deficient in Crohn's disease. *Gut* **65**, 415–425 (2016).
- Louis, P. et al. Restricted distribution of the butyrate kinase pathway among butyrate-producing bacteria from the human colon. *J. Bacteriol.* **186**, 2099–2106 (2004).
- Litvak, Y., Byndloss, M. X. & Bäuml, A. J. Colonocyte metabolism shapes the gut microbiota. *Science* **362**, eaat9076 (2018).
- Kriss, M., Hazleton, K. Z., Nusbacher, N. M., Martin, C. G. & Lozupone, C. A. Low diversity gut microbiota dysbiosis: drivers, functional implications and recovery. *Curr. Opin. Microbiol.* **44**, 34–40 (2018).
- Ding, T. & Schloss, P. D. Dynamics and associations of microbial community types across the human body. *Nature* **509**, 357–360 (2014).

20. Gardiner, B. J. et al. Clinical and microbiological characteristics of *Eggerthella lenta* bacteremia. *J. Clin. Microbiol.* **53**, 626–635 (2015).
21. Mazmanian, S. K., Liu, C. H., Tzianabos, A. O. & Kasper, D. L. An immunomodulatory molecule of symbiotic bacteria directs maturation of the host immune system. *Cell* **122**, 107–118 (2005).
22. Wexler, H. M. Bacteroides: the good, the bad, and the nitty-gritty. *Clin. Microbiol. Rev.* **20**, 593–621 (2007).
23. Ridker, P. M. et al. Reduction in C-reactive protein and LDL cholesterol and cardiovascular event rates after initiation of rosuvastatin: a prospective study of the JUPITER trial. *Lancet* **373**, 1175–1182 (2009).
24. Muscogiuri, G. et al. The good and bad effects of statins on insulin sensitivity and secretion. *Endocr. Res.* **39**, 137–143 (2014).
25. Khan, T. J. et al. Effect of atorvastatin on the gut microbiota of high fat diet-induced hypercholesterolemic rats. *Sci. Rep.* **8**, 662 (2018).
26. Maier, L. et al. Extensive impact of non-antibiotic drugs on human gut bacteria. *Nature* **555**, 623–628 (2018).
27. Forslund, K. et al. Disentangling type 2 diabetes and metformin treatment signatures in the human gut microbiota. *Nature* **528**, 262–266 (2015).
28. Zeiser, R. Immune modulatory effects of statins. *Immunology* **154**, 69–75 (2018).
29. Grip, O., Janciauskiene, S. & Bredberg, A. Use of atorvastatin as an anti-inflammatory treatment in Crohn's disease. *Br. J. Pharmacol.* **155**, 1085–1092 (2008).
30. Ungaro, R. et al. Statins associated with decreased risk of new onset inflammatory bowel disease. *Am. J. Gastroenterol.* **111**, 1416–1423 (2016).

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

© The Author(s), under exclusive licence to Springer Nature Limited 2020

¹Laboratory of Molecular Bacteriology, Department of Microbiology and Immunology, Rega Institute, KU Leuven, Leuven, Belgium. ²Center for Microbiology, VIB, Leuven, Belgium. ³Nutrition and Obesity: Systemic Approaches Research Unit (NutriOmics), INSERM, Sorbonne Université, Paris, France. ⁴Institute of Cardiometabolism and Nutrition, Integromics Unit, Paris, France. ⁵Novo Nordisk Foundation Center for Basic Metabolic Research, Faculty of Health and Medical Sciences, University of Copenhagen, Copenhagen, Denmark. ⁶Nutrition Department, Pitie-Salpêtrière Hospital, Assistance Publique Hôpitaux de Paris, Paris, France. ⁷Medical Department III – Endocrinology, Nephrology, Rheumatology, University of Leipzig Medical Center, Leipzig, Germany. ⁸Experimental and Clinical Research Center, Charité-Universitätsmedizin and Max-Delbrück Center, Berlin, Germany. ⁹Structural and Computational Biology, European Molecular Biology Laboratory, Heidelberg, Germany. ¹⁰Max Delbrück Center for Molecular Medicine in the Helmholtz Association (MDC), Berlin, Germany. ¹¹DZHK (German Centre for Cardiovascular Research), Partner Site Berlin, Berlin, Germany. ¹²Unité de Modélisation Mathématique et Informatique des Systèmes Complexes, UMMISCO, Sorbonne Université, IRD, Bondy, France. ¹³Wallenberg Laboratory, Department of Molecular and Clinical Medicine and Sahlgrenska Academy, University of Gothenburg, Gothenburg, Sweden. ¹⁴Université Paris-Saclay, INRAE, Metagenopolis, Jouy en Josas, France. ¹⁵Diabetes Department, Pitie-Salpêtrière Hospital, Assistance Publique Hôpitaux de Paris, Paris, France. ¹⁶UF Biomarqueurs Inflammatoires et Métaboliques, Biochemistry and Hormonology Department, Tenon Hospital, Assistance Publique Hôpitaux de Paris, Paris,

France. ¹⁷Centre de Recherche Saint-Antoine, Sorbonne Université-INSERM UMR-S 938, IHU ICAN, Paris, France. ¹⁸Institute of Science and Technology for Brain-Inspired Intelligence, Fudan University, Shanghai, China. ¹⁹NICO Cardio-oncology Program, CIC-1421, Department of Pharmacology, Pitie-Salpêtrière Hospital, Assistance Publique Hôpitaux de Paris, INSERM, Sorbonne Université, Paris, France. ²⁰Université de Paris, PARCC, INSERM, Paris, France. ²¹CIC1418 and DMU CARTE, Assistance Publique-Hôpitaux de Paris, Hôpital Européen Georges-Pompidou, Paris, France. ²²Department of Cardiology, Rigshospitalet, University of Copenhagen, Copenhagen, Denmark. ²³Computational and Systems Medicine, Department of Metabolism, Digestion and Reproduction, Faculty of Medicine, Imperial College London, London, UK. ²⁴Genomic and Environmental Medicine, National Heart & Lung Institute, Faculty of Medicine, Imperial College London, London, UK. ²⁵Sorbonne Paris Cité Epidemiology and Statistics Research Centre (CRESS), U1153 INSERM, U1125, INRA, CNAM, University of Paris, Nutritional Epidemiology Research Team (EREN), Bobigny, France. ²⁶Department of Clinical Biochemistry, Rigshospitalet, University of Copenhagen, Copenhagen, Denmark. ²⁷Biocyte Solutions, Heidelberg, Germany. ²⁸Department of Biology and Biological Engineering, Chalmers University of Technology, Gothenburg, Sweden. ²⁹Helmholtz Institute for Metabolic, Obesity and Vascular Research (HI-MAG) of the Helmholtz Zentrum München at the University of Leipzig, Leipzig, Germany. ³⁰Molecular Medicine Partnership Unit, University of Heidelberg and European Molecular Biology Laboratory, Heidelberg, Germany. ³¹Department of Bioinformatics, Biocenter, University of Würzburg, Würzburg, Germany. ³⁶These authors contributed equally: Sara Vieira-Silva, Gwen Falony, Eugeni Belda. ³⁷These authors jointly supervised this work: Karine Clément, Jeroen Raes. *A list of participants and their affiliations appears at the end of the paper. ^{32a}e-mail: karine.clement@inserm.fr; jeroen.raes@kuleuven.be

MetaCardis Consortium

Renato Alves⁹, Chloe Amouyal^{3,15}, Ehm Astrid Andersson Galijatovic⁵, Olivier Barthelemy³², Jean-Paul Batisse³², Magalie Berland¹⁴, Randa Bittar³³, Hervé Blottière¹⁴, Frederic Bosquet¹⁵, Rachid Boubrit³², Olivier Bourron¹⁵, Mickael Camus¹⁴, Dominique Cassuto⁶, Cecile Ciangura^{6,15}, Jean-Philippe Collet³², Maria-Carlota Dao³, Jean Debedat³, Morad Djebbar³², Angélique Doré¹⁴, Line Engelbrechtsen⁵, Soraya Fellahi^{16,17}, Sebastien Fromentin¹⁴, Philippe Giral¹⁴, Marianne Graine⁴, Agnes Hartemann¹⁵, Bolette Hartmann⁵, Gerard Helft³², Serge Hercberg²⁵, Malene Hornbak³, Richard Isnard³², Sophie Jaqueminet¹⁵, Niklas Rye Jørgensen³⁵, Hanna Julienne¹⁴, Johanne Justesen⁵, Judith Kammer⁷, Mathieu Kerneis³², Jean Khemis⁶, Nikolaj Krarup⁵, Michael Kuhn⁹, Aurélie Lampuré³, Véronique Lejard¹⁴, Florence Levenez¹⁴, Lea Lucas-Martini⁹, Robin Massey¹⁴, Nicolas Maziers¹⁴, Jonathan Medina-Stamminger⁶, Lucas Moitinho-Silva⁹, Gilles Montalescot³², Sandrine Moutel⁶, Laetitia Pasero Le Pavin¹⁴, Christine Poitou-Bernert^{3,6}, Françoise Pousset³², Laurence Pouzoulet³⁴, Sebastian Schmidt⁹, Johanne Silvain³², Mathilde Svendstrup⁵, Timothy Swartz³⁴, Thierry Vanduyvenboden¹⁴, Camille Vatie⁶, Eric Verger³ & Stefanie Walther⁷

³²Cardiology Department, Pitie-Salpêtrière Hospital, Assistance Publique Hôpitaux de Paris, Paris, France. ³³Biochemistry Department, Pitie-Salpêtrière Hospital, Assistance Publique Hôpitaux de Paris, Paris, France. ³⁴Endocrinology Department, Pitie-Salpêtrière Hospital, Assistance Publique Hôpitaux de Paris, Paris, France. ³⁵Department of Clinical Biochemistry, Rigshospitalet, Glostrup University of Copenhagen, Copenhagen, Denmark.

Methods

Data reporting

No statistical methods were used to predetermine sample size. The experiments were not randomized and the investigators were not blinded to allocation during experiments and outcome assessment.

Sample collection

Ethical compliance. Ethical approval was obtained from the Ethics Committee CPP Ile-de France, Ethics Committee at the Medical Faculty at the University of Leipzig, and the Ethical Committees of the Capital Region of Denmark. The study protocol (also comprising an interventional arm which is not part of the analysis presented) was registered at <https://clinicaltrials.gov> (study number NCT02059538). The study design (observational cohort study) complied with all relevant ethical regulations, aligning with the Helsinki Declaration and in accordance with European privacy legislation. All participants provided written informed consent.

Study cohort. The $n = 888$ transnational Body Mass Index spectrum (BMIS) cohort was assembled as part of the overall MetaCardis recruitment efforts (Supplementary Fig. 1). Participants were recruited between 2013 and 2015 in the clinical departments of the Pitié-Salpêtrière Hospital (Paris, France), the Integrated Research and Treatment Center for Adiposity Diseases (Leipzig, Germany), and in the Novo Nordisk Foundation Center for Basic Metabolic Research (Copenhagen, Denmark). Potential participants were evaluated for suitability according to standardized inclusion and exclusion criteria across the three recruitment centres. Exclusion criteria included history of abdominal cancer/radiation therapy on the abdomen, history of intestinal resection (except for appendectomy), acute or chronic inflammatory or infectious diseases (including hepatitis C virus, hepatitis B virus and HIV), history of organ transplantation or receiving immunosuppressive therapy, severe kidney failure (MDRD glomerular filtration rate < 50 ml (min 1.73m^2) $^{-1}$), or drug or alcohol addiction. All study participants had to be free of any antibiotic use in the three months before inclusion. The BMIS ($n = 888$) cohort consisted of a MetaCardis sub-cohort, defined by exclusion of cardiovascular patients (defined in the MetaCardis consortium study protocol as patient groups 4, 5, 6 and 7) and any individual with type-2 diabetes. Diagnosis of type-2 diabetes was defined using the American Diabetes Association definition: fasting glycemia > 6.9 mmol l^{-1} and/or 2 h values in the oral glucose tolerance test > 11 mmol l^{-1} and/or haemoglobin A1c (HbA1c, glycated haemoglobin) $\geq 6.5\%$ and/or use of any antidiabetic treatment. The MetaCardis project sample size calculation was focused on the objectives of multi-omics integration and metagenomic-wide metabolome-wide association study (MW2AS) across groups of patients ranging different cardiometabolic phenotypes. On the basis of unpublished data from consortium partners, a sample size of 2,000 individuals was deemed required to detect a significant association (with and without concomitant risk factors). No specific sample size calculation was performed before BMIS sub-cohort recruitment. On the basis of the baseline prevalence of Bact2 enterotype (with baseline defined as lean/overweight individuals $P(\text{Bact2}) = 14\%$) in the amplicon-sequenced FGFP cohort, the present study cohort size enabled us to identify a minimum difference of 7.4% in Bact2 prevalence between the two groups: lean or overweight ($n = 414$) versus obese ($n = 474$) as significant (power = 80%, $\alpha = 0.05$).

Validation cohorts. MetaCardis cardiovascular disease (CVD, $n = 282$). The CVD cohort was recruited as described above as part of the MetaCardis cohort, and corresponds to patients with cardiovascular disease and without diabetes, defined in the MetaCardis consortium study protocol as patient groups 4, 5, 6 and 7. Flemish Gut Flora Project (FGFP, $n = 2,345$). The FGFP cohort is part of a population-level cross-sectional sampling of the Flemish population described in

ref.¹⁰ and re-sequenced with dual-indexed HiSeq amplicon sequencing as analysed in ref.³¹. Ethical approval for the FGFP sampling was granted by the Commissie Medische Ethiek UZ-VUB (B.U.N.143201215505) and the Ethische Commissie Onderzoek UZ/KU Leuven (S58125). The inclusion and exclusion criteria defined for recruitment of the MetaCardis cohort and, more specifically, the BMIS subset, were applied to the FGFP: inclusion age between 18 and 75 years old, exclusion of acute or chronic inflammatory or infectious diseases (notably diagnosis of inflammatory bowel disease and recent gastroenteritis), and exclusion of patients with diabetes—defined as having a diagnosis of diabetes or increased glycated haemoglobin A1c levels (HbA1c $\geq 6.5\%$), or the use of any antidiabetic treatment. The disease diagnoses used for exclusion were reported by the general practitioners of the participants. The medical questionnaire and blood sampling for analysis (including HbA1c) were performed within one week of faecal sampling.

Sample collection. Faeces were collected according to International Human Microbiome Standards (IHMS) guidelines (modified SOP 04 V1 (collection without anaerobic bag)). In brief, participants were handed a collection kit, collected samples at home, and stored them temporarily (less than 48 h) at -20°C until they were transported frozen (on dry ice) to the collection centre (Pitié-Salpêtrière Hospital (France), University Hospital of Leipzig (Germany) or Frederiksberg Hospital (Denmark)). Blood samples were collected during the clinical examination visit after overnight fasting.

Metadata collection. Participant phenotyping was performed according to standardized operational procedures and included the acquisition of biological samples and the assessment of clinical parameters and anthropometrics including age, gender, smoking status, weight, height, BMI, blood pressure, body composition, and waist and hip circumference measurements. Body fat mass and fat-free mass were determined through bioelectrical impedance analysis. Systolic and diastolic blood pressure were measured using a mercury sphygmomanometer (measures were taken three times on each arm; the mean of the last two measurements on the right arm was used for analyses). During the interview at the clinical visit, a detailed list of prescribed medications (based on direct recall or medication list when provided) as well as the medical history of the patient was compiled. Subjects were questioned on adherence to their medication plan. Five-year antibiotic intake was assessed by recall in France and Denmark, whereas participants in Germany were requested to provide medication anamnesis from their general practitioners or physicians (drugs prescribed over the past five years). All medication data was curated jointly by the study physicians at each centre so as to harmonize presentation. The metadata necessary for reproducing the results presented in the article are available in Supplementary Table 2.

Sample analyses

Blood analyses. Blood metabolic markers were assessed in local routine laboratories. Analyses of adipokines, measures of glycaemia, inflammatory markers, and free fatty acids were centralized; plasma and serum samples were stored at the respective clinical centres at -80°C until shipment to a central measuring facility. Blood cell counts (leukocytes, monocytes, neutrophils and immune cells) were measured using flow cytometry as described previously³². Fasting glucose, total cholesterol, high-density-lipoprotein cholesterol, triglycerides and HbA1c were measured using enzymatic methods. LDL-cholesterol concentrations were measured enzymatically for German participants; values for French and Danish subjects were calculated using the Friedwald equation. Kinetic assays based on coupled enzyme systems were used to measure alanine aminotransferase, aspartate aminotransferase and γ -glutamyltransferase levels. Free fatty acid concentrations were assessed by photometrics (Diasys Diagnostic Systems). A chemiluminescence assay (Insulin Architect, Abbott) was used to measure

serum insulin and C-peptide levels in a fasting state and at 30 and 120 min during an oral glucose tolerance test. Serum leptin was determined using the Human Leptin Quantikine ELISA Kit (R&D Systems); adiponectin was measured using an ELISA sandwich assay (HMW & Total Adiponectin ELISA Kit, ALPCO). Levels of hsCRP were determined by an IMMAGE automatic immunoassay system (Beckman-Coulter). Blood concentrations of high-sensitivity interleukin 6 (hsIL6) and CD14 were measured using the Human IL-6 Quantikine HS and the Human Quantikine ELISA Kit (R&D Systems), respectively. A Luminex assay (ProcartaPlex Mix&Match Human 13-plex, eBioscience) was set up to measure the following cytokines: interferon gamma-induced protein 10 (IP-10), C-X-C motif chemokine ligand 5 (CXCL5), CC-Chemokine ligand 2 (CCL2), Eotaxine, Interleukine 7 (IL-7), macrophage migration inhibitory factor (MIF), macrophage inflammatory protein 1 β (MIP 1 β), stromal cell-derived factor 1 (SDF1) and vascular endothelial growth factor A (VEGFA).

Metagenomic analyses of faecal samples. Total faecal DNA was extracted following the International Human Microbiome Standards (IHMS) guidelines (SOP 07 V2 H) and sequenced using an Ion proton system (Thermo Fisher Scientific) resulting in 23.3 ± 4.0 million (mean \pm s.d.) 150-bp single-end reads per sample on average. Reads were cleaned using AlienTrimmer (v0.2.4)³³ to remove resilient sequencing adapters and to trim low quality nucleotides at the 3' side (quality and length cut-off of 20 and 45 bp, respectively). Cleaned reads were subsequently filtered from human and potential food contaminant DNA (using human genome RCh37-p10, *Bos taurus* and *Arabidopsis thaliana* with an identity score threshold of 97%). Gene abundance profiling was performed using the 9.9-million-gene integrated reference catalogue of the human microbiome³⁴. Filtered high-quality reads were mapped with an identity threshold of 95% to the 9.9-million-gene catalogue using BowTie (v.2.2.6) included in the METEOR software³⁵. A gene abundance table was generated by means of a two-step procedure using METEOR. First, the uniquely mapping reads (reads mapping to a single gene in the catalogue) were attributed to their corresponding genes. Second, shared reads (reads that mapped with the same alignment score to multiple genes) were attributed according to the ratio of their unique mapping counts. The gene abundance table was processed for rarefaction and normalization and further analysis using the R package MetaOMineR³⁶. To decrease technical bias due to different sequencing depth and avoid any artefacts of sample size on low-abundance genes, read counts were rarefied. The gene abundance table was rarefied to 10 million reads per sample by random sampling of 10 million mapped reads without replacement. The resulting rarefied gene abundance table was normalized according to the FPKM (fragments per kilobase of transcript per million mapped reads) strategy (normalization by the gene size and the number of total mapped reads reported in frequency) to give the gene abundance profile table and binned by functional and phylogenetic categories as carried out within the MOCAT2 framework³⁷. 1,436 metagenomic species (MGS; co-abundant gene groups with more than 500 genes corresponding to microbial species) were clustered from 1,267 human gut metagenomes used to construct the 9.9-million-gene catalogue³⁴, as described previously³⁸. MGS abundances were estimated as the mean abundance of the 50 genes defining a robust centroid of the cluster (if more than 10% these genes gave positive signals). MGS taxonomical annotation was performed using all genes by sequence similarity using NCBI blast N; a species-level assignment was given if more than 50% of the genes matched the same reference genome of the NCBI database (November 2016 version) at a threshold of 95% of identity and 90% of gene length coverage. The remaining MGS were assigned to a given taxonomical level from genus to superkingdom if more than 50% of their genes had the same level of assignment. Microbial gene richness (gene count) was calculated by counting the number of genes that were detected

at least once in a given sample, using the average number of genes counted in ten independent rarefaction experiments.

Determination of faecal microbial load. Microbial loads of faecal samples of were determined as described previously^{1,2}. In brief, 0.2 g frozen (-80°C) aliquots were dissolved in physiological solution ($9\text{ g l}^{-1}\text{ NaCl}$; Baxter S.A.) to a total volume of 100 ml. Subsequently, the slurry was diluted 1,000 times. Samples were filtered using a sterile syringe filter (pore size $5\text{ }\mu\text{m}$; Sartorius Stedim Biotech). Next, 1 ml of the microbial cell suspension obtained was stained with $1\text{ }\mu\text{l}$ SYBR Green I (1:100 dilution in DMSO; shaded 15 min incubation at 37°C ; 10,000 concentrate, Thermo Fisher Scientific). The flow cytometry analysis was performed using a C6 Accuri flow cytometer (BD Biosciences)³⁹. Fluorescence events were monitored using the FL1 533/30 nm and FL3 $> 670\text{ nm}$ optical detectors. In addition, forward and sideward-scattered light was also collected. The BD Accuri CFlow (v.1.0.264.21) software was used to gate and separate the microbial fluorescence events on the FL1/FL3 density plot from the faecal sample background. A threshold value of 2,000 was applied on the FL1 channel. The gated fluorescence events were evaluated on the forward/sideward density plot, so as to exclude remaining background events. Instrument and gating settings were kept identical for all samples (fixed staining/gating strategy³⁹; Supplementary Fig. 2). On the basis of the exact weight of the aliquots analysed, cell counts were converted to microbial loads per gram of faecal material.

Analyses of faecal metagenomes

Quantitative microbiome profiling. Phylogenetic quantitative microbiome profiles were built using a modified version of the pipeline described in ref. ¹. In short, sample abundance profiles were downsized to even sampling depth, defined as the ratio between sampling size (average mOTU marker genes coverage⁴⁰) and microbial load (average total cell count per gram of frozen faecal material). The sequencing depth of each sample was rarefied to the level necessary to equate the minimum observed sampling depth in the cohort. The rarefied mOTU abundance matrix was converted into numbers of cells per gram and quantitative microbiome profiling matrices created for phylum to species levels. Functional quantitative microbiome profiles and quantitative co-abundance gene groups³⁸ profiles were constructed by multiplication of relative proportions to an indexing factor proportional to the microbial cell densities of the samples (load), defined as the sample load divided by the median load over the entire MetaCardis cohort. The processed microbiome profiles can be downloaded at <http://raeslab.org/software/BMIS/>.

Customized module analyses. Customized module sets included previously described gut metabolic modules¹¹ covering bacterial and archaeal metabolism specific to the human gut environment with a focus on anaerobic fermentation processes, expanded with a specific set of six modules focusing on bacterial trimethylamine metabolism⁴¹. Additionally, following a previously published strategy to build manually curated gut-specific metabolic modules^{11,31}, we constructed a new set of modules to describe and map microbial phenylpropanoid metabolism (phenylpropanoid metabolism modules, PPM) from shotgun metagenomic data. This set of 20 modules, following KEGG syntax, is provided in the Supplementary Information, including references to the original publications in which the pathways were described (Supplementary Table 3). Abundances of customized modules were derived from the orthologue abundance tables using Omixer-RPM v1.0 (<https://github.com/raeslab/omixer-rpm>)^{11,42}. The coverage of each metabolic variant encoded in a module was calculated as the number of steps for which at least one of the orthologous groups was found in a metagenome, divided by the total number of steps constituting the variant. The presence or absence of a module was identified with a detection threshold of more than 66% coverage to provide tolerance to

Article

misannotations and missing data in metagenomes. Module abundance was calculated as the median of orthologue abundances in the pathway with maximum coverage.

Statistical analyses

Statistical analyses were performed in R using the following packages: *vegan*⁴³ v.2.5-3, *phyloseq*⁴⁴ v.1.26.0, *FSA*⁴⁵ v.0.8.24, *coin*⁴⁶ v.1.2-2, *DirichletMultinomial*⁴⁷ v.1.24.0, *Hmisc*⁴⁸ v.4.1-1, *car*⁴⁹ v.3.0-2, *sjstats*⁵⁰ v.0.17.5, and *nnet*⁵¹ v.7.3-12. All statistical tests used were two-sided. All *P* values were corrected for multiple testing when appropriate using the Benjamini–Hochberg method (P_{adj}), only $P_{adj} < 0.05$ were reported as significant.

Faecal microbiome derived features and visualization. Observed richness was calculated using *phyloseq*⁴⁴. Microbiome inter-individual variation was visualized by principal coordinates analysis using Bray–Curtis dissimilarity on the genus-level relative abundance matrix with Hellinger transformation.

Partitioning of microbiome variation across clinical explanatory variables. The estimation of the explanatory power of clinical features regarding relative, genus-level, microbiome profiles variation was performed using univariate or multivariate stepwise distance-based redundancy analysis as implemented in the R package *vegan*⁴³.

Microbiome community typing. Enterotyping (or community typing) of the genus-level abundance microbial profiles with Hellinger transformation was performed on the basis of the Dirichlet multinomial mixtures (DMM) approach implemented in the R package *DirichletMultinomial*, as described in ref.⁵² on the whole of the $n = 2,022$ MetaCardis cohort. Although the dissimilarity/distance-based approaches were applied to screen for covariate-associated microbiome trends throughout the whole of the BMIS cohort, DMM-based stratification allows identification of covariates not only associated with the strata, but also linked to fluctuations in the prevalence of one (or more) particular microbiota constellation(s). This makes enterotyping a valuable strategy when assessing microbiome variation in pathologies that are not expected to be characterized by generalized dysbiosis with varying severity according to diagnosis⁵³, but—by contrast—by the increased occurrence of a single dysbiotic community type with prevalence depending on the condition studied^{1,2,31}, as proposed here for obesity.

Microbiome features and clinical features associations. Taxa unclassified at the genus level or present in fewer than 20% of samples were excluded from the statistical analyses. Pearson or Spearman correlations were used, respectively, for linear or rank-order correlations between continuous variables, including genera abundances and metadata. The Mann–Whitney *U*-test was used to test median differences of continuous variables between two different groups. For more than two groups, the Kruskal–Wallis test with post-hoc Dunn test were used. Statistical differences in the prevalence of enterotypes between groups were evaluated using pairwise Fisher's exact tests. Modelling the association between the prevalence of enterotypes (Bact1, Bact2, Prev, Rum) or Bact2 prevalence (Bact2 = Yes/No) and single (univariate) or multiple (multivariate) dependent variables (clinical metadata features) was performed using generalized linear models, namely multinomial or binomial logistic regression (for enterotypes or Bact2 prevalences, respectively) with significance evaluated by likelihood ratio tests using the R package *car*. Risk ratio estimates (and their confidence intervals) were retrieved using the R package *sjstats*, by conversion of the odds ratios of the generalized linear models⁵⁴, the latter corresponding to exponential transformation of the model coefficients.

Reporting summary

Further information on research design is available in the Nature Research Reporting Summary linked to this paper.

Data availability

Raw amplicon sequencing data used in this study have been deposited in the EMBL-EBI European Nucleotide Archive (ENA) under accession number PRJEB37249. The metadata and processed microbiome data required for the reanalysis of results presented in the manuscript are respectively provided as Supplementary Table 2 and available for download at <http://raeslab.org/software/BMIS/>. For clinical cohort-related questions, contact K.C.

- Valles-Colomer, M. et al. The neuroactive potential of the human gut microbiota in quality of life and depression. *Nat. Microbiol.* **4**, 623–632 (2019).
- Touch, S. et al. Mucosal-associated invariant T (MAIT) cells are depleted and prone to apoptosis in cardiometabolic disorders. *FASEB J.* **32**, 5078–5089 (2018).
- Criscuolo, A. & Brisse, S. AlienTrimmer: a tool to quickly and accurately trim off multiple short contaminant sequences from high-throughput sequencing reads. *Genomics* **102**, 500–506 (2013).
- Li, J. et al. An integrated catalog of reference genes in the human gut microbiome. *Nat. Biotechnol.* **32**, 834–841 (2014).
- Cotillard, A. et al. Dietary intervention impact on gut microbial gene richness. *Nature* **500**, 585–588 (2013).
- Prifti, E. & Le Chatelier, E. MetaOMineR: a quantitative metagenomics data analyses pipeline. R package v.1.1 (2015).
- Kultima, J. R. et al. MOCAT2: a metagenomic assembly, annotation and profiling framework. *Bioinformatics* **32**, 2520–2523 (2016).
- Nielsen, H. B. et al. Identification and assembly of genomes and genetic elements in complex metagenomic samples without using reference genomes. *Nat. Biotechnol.* **32**, 822–828 (2014).
- Prest, E. I., Hammes, F., Köttsch, S., van Loosdrecht, M. C. M. & Vrouwenvelder, J. S. Monitoring microbiological changes in drinking water systems using a fast and reproducible flow cytometric method. *Water Res.* **47**, 7131–7142 (2013).
- Kultima, J. R. et al. MOCAT: a metagenomics assembly and gene prediction toolkit. *PLoS ONE* **7**, e47656 (2012).
- Falony, G., Vieira-Silva, S. & Raes, J. Microbiology meets big data: the case of gut microbiota-derived trimethylamine. *Annu. Rev. Microbiol.* **69**, 305–321 (2015).
- Darzi, Y., Falony, G., Vieira-Silva, S. & Raes, J. Towards biome-specific analysis of meta-omics data. *ISME J.* **10**, 1025–1028 (2016).
- Oksanen, J. et al. *vegan: Community Ecology Package*. R package v.2.2-1 (2015).
- McMurdie, P. J. & Holmes, S. *phyloseq: an R package for reproducible interactive analysis and graphics of microbiome census data*. *PLoS ONE* **8**, e61217 (2013).
- Ogle, D. H. FSA: Fisheries Stock Analysis. R package v.0.8.13. (2017).
- Hothorn, T., Hornik, K., van de Wiel, M. A. & Zeileis, A. A Lego system for conditional inference. *Am. Stat.* **60**, 257–263 (2006).
- Morgan, M. *DirichletMultinomial: Dirichlet-multinomial mixture model machine learning for microbiome data*. R package v.1.18.0 (2017).
- Harrell, F. E. *Hmisc: Harrell Miscellaneous*. R package v.4.1-1 (2018).
- Fox, J. & Weisberg, S. *An R Companion to Applied Regression*. (Sage, 2011).
- Lüdtke, D. *sjstats: Statistical Functions for Regression Models* v.0.17.5 (2019).
- Venables, W. N. & Ripley, B. D. *Modern Applied Statistics with S*. (Springer, 2002).
- Holmes, I., Harris, K. & Quince, C. Dirichlet multinomial mixtures: generative models for microbial metagenomics. *PLoS ONE* **7**, e30126 (2012).
- Duvallet, C., Gibbons, S. M., Gurry, T., Irizarry, R. A. & Alm, E. J. Meta-analysis of gut microbiome studies identifies disease-specific and shared responses. *Nat. Commun.* **8**, 1784 (2017).
- Grant, R. L. Converting an odds ratio to a range of plausible relative risks for better communication of research findings. *Br. Med. J.* **348**, f7450 (2014).

Acknowledgements We thank the study participants and nurses for their contributions to the project. MetaCardis was funded by European Union's Seventh Framework Programme for research, technological development and demonstration under grant agreement HEALTH-F4-2012-305312 (MetaCardis project) and the French National Agency of Research (ANR; 'Investissement d'Avenir' FORCE, Metagenopolis grant ANR-11-DPBS-0001 and ICAN ANR-10-IAHU-05). The promotor of the clinical study was the Assistance Publique Hôpitaux de Paris (APHP). S.V.-S. was supported by a post-doctoral fellowship from the Research Foundation Flanders (FWO Vlaanderen). The Raes laboratory is supported by the VIB Grand Challenges programme, KU Leuven, the Rega Institute for Medical Research, and the FWO EOS program (30770923). The Novo Nordisk Foundation Center for Basic Metabolic Research is an independent research institution at the University of Copenhagen partially funded by an unrestricted donation from the Novo Nordisk Foundation. M.-E.D. was funded by the NIHR Imperial Biomedical Research Centre.

Author contributions M.-E.D., S.D.E., P.G., J.P.G., T.H., J.J.H., L.K., I.L., J.N., J.-M.O., M.S., H.V., J.-D.Z., P.B., O.P., F.B., K.C. (the MetaCardis Consortium coordinator) and J.R. conceived the MetaCardis study protocol, including clinical standard operating procedures, study objectives and study design. T.N., J.A.-W. and R.C. coordinated recruitment and sample collection efforts over the different cohorts. T.N., J.A.-W., R.C. and K.A. curated and harmonized the clinical metadata. S.V.-S., G.F., E.B., T.N., J.A.-W., S.K.F., K.A., R.C., M.V.-C., S.P., E.P., V.T., N.P., E.L.C., F.A.,

J.-P.B., L.P.C., N.G., T.H.H., J.-S.H., C.L., H.K.P., B.Q., C.R., H.R., J.-E.S., N.B.S., S.T. and the MetaCardis Consortium assisted in sample collection, analyses, and/or data pre-processing and exploration. Faecal microbial DNA extraction and shotgun sequencing was performed by N.P., E.L.C. and S.F. Flow-cytometry-based faecal microbial load estimations were performed by T.T.D.N. Statistical analyses were designed and executed by S.V.-S., G.F., E.B., K.A., S.K.F. and M.V.-C. The manuscript was drafted by G.F., S.V.-S., K.C. and J.R. All authors revised the article and approved the final version for publication.

Competing interests J.R., S.V.-S., G.F. and M.V.-C. are listed as inventors on patent application PCT/EP2018/084920, in the name of VIB VZW, Katholieke Universiteit Leuven, KU Leuven R&D

and Vrije Universiteit Brussel, covering the features of the microbiome associated with inflammation described in ref. ².

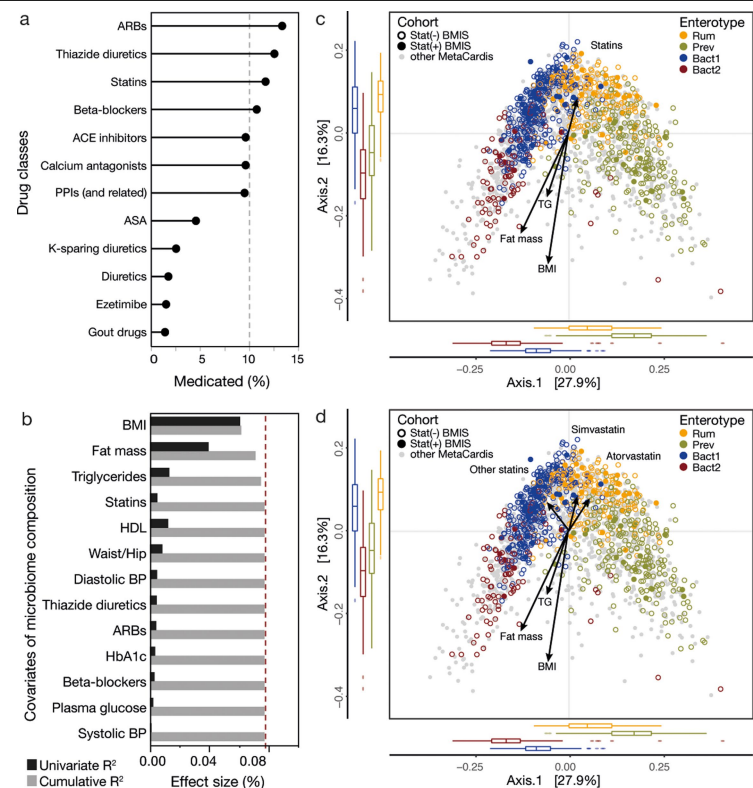
Additional information

Supplementary information is available for this paper at <https://doi.org/10.1038/s41586-020-2269-x>.

Correspondence and requests for materials should be addressed to K.C. or J.R.

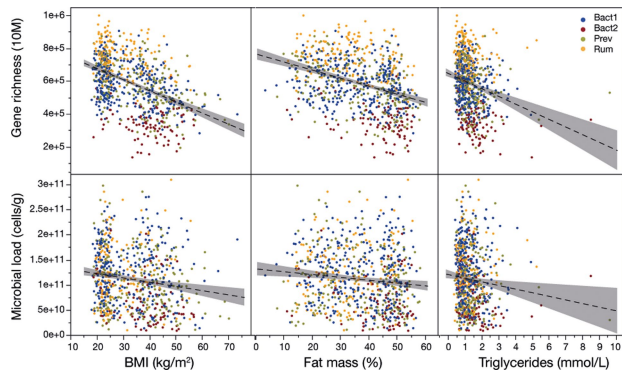
Peer review information *Nature* thanks Peter Turnbaugh and the other, anonymous, reviewer(s) for their contribution to the peer review of this work.

Reprints and permissions information is available at <http://www.nature.com/reprints>.

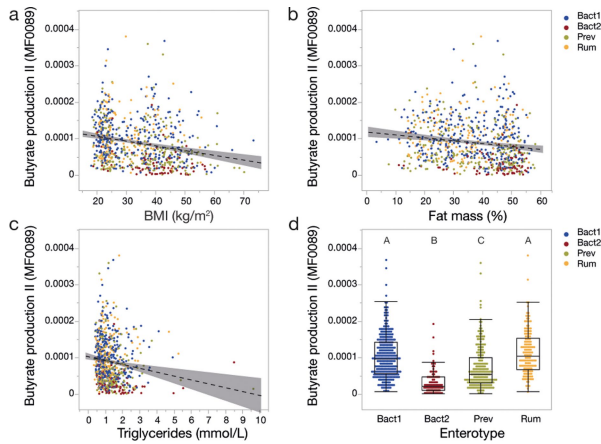


Extended Data Fig. 1 | Microbiome variation in the BMIS cohort ($n = 888$ participants). **a**, Percentage of subjects in the BMIS cohort taking medication of the stated drug classes. ACE inhibitors, angiotensin converting enzyme inhibitors; ARB, angiotensin II receptor blockers; ASA, acetylsalicylic acid; PPI, proton-pump inhibitors. **b**, Best model explaining inter-individual microbiome variation based on obesity-defining and metabolic-syndrome-defining variables as well as on most frequently disclosed therapeutics (taken by more than 10% of participants; Supplementary Table 4). Explanatory power of the included variables are reported for the variables taken individually (black bars; $n = 888$ biologically independent samples, univariate dbrDA) or in a multivariate model (grey bars; $n = 888$ biologically independent samples, multivariate dbrDA). **c**, Principal coordinates analysis of inter-individual differences (genus level Bray–Curtis dissimilarity) in the microbiome profiles

of the BMIS cohort ($n = 888$ biologically independent samples, data points coloured by enterotypes (Extended Data Fig. 4)) with the rest of the MetaCardis dataset in the background ($n = 1,134$, grey dots). Full and open circles corresponding to statin-medicated (Stat(+)) and non-statin-medicated participants (Stat(-)), respectively. Arrows represent the effect sizes of a post hoc fit of significant microbiome covariates identified in the multivariate model in **b**. **d**, Same principal coordinates analysis as in **c**, with the statin intake variable split into the separate statin classes ($n = 888$ biologically independent samples, simvastatin ($n = 51$), atorvastatin ($n = 33$) and other statins ($n = 22$); Supplementary Table 4). In **c**, **d**, the body of the box plot represents the first and third quartiles of the distribution, the line represents the median, and the whiskers extend from the quartiles to the last data point within $1.5 \times$ the interquartile range (IQR), with outliers beyond.

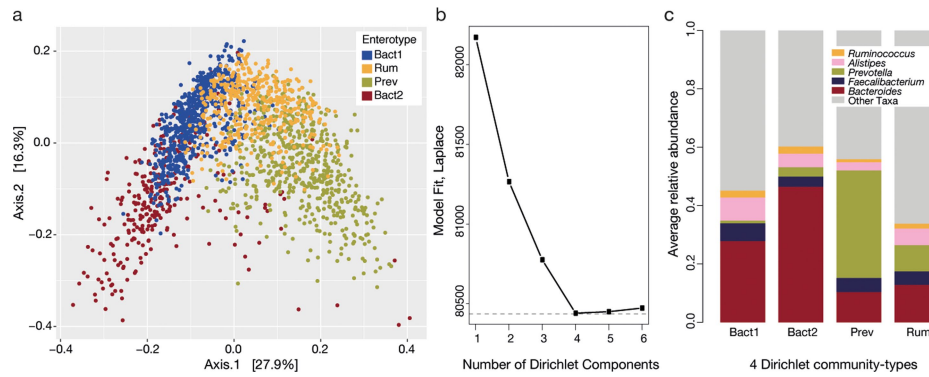


Extended Data Fig. 2 | The association of BMI, fat mass percentage and serum fasting triglyceride levels with faecal microbial gene richness and faecal microbial load in the non-statin-medicated BMIS cohort ($n = 782$ participants). All three covariates were found to be associated with both microbiome gene richness ($n = 711$ biologically independent samples, Spearman's $\rho = -0.45$ to -0.26 , $P_{\text{adj}} = 4.0 \times 10^{-39}$ to 1.6×10^{-13}), a proxy for microbial biodiversity previously suggested as a marker of metabolic health in obese individuals⁸, and faecal microbial load ($n = 711$ biologically independent samples, Spearman's $\rho = -0.17$ to -0.13 , $P_{\text{adj}} = 4.1 \times 10^{-6}$ to 3.1×10^{-4} ; Supplementary Table 7). Adjustment for multiple testing (P_{adj}) was performed using the Benjamini–Hochberg method. Least square linear regression lines (dashed line) with 95% confidence interval (grey shading) are provided for visual representation of the non-parametric testing provided in Supplementary Table 7. Data points are coloured by enterotype classification.



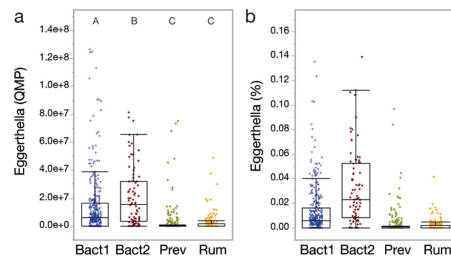
Extended Data Fig. 3 | Association between the variation in quantitative butyrate production potential and the BMI, fat mass percentage and triglycerides levels of participants, or the enterotype classification of the samples, in the non-statin-medicated BMIS cohort ($n = 782$ participants).

Quantitative functional microbiome profiles were constructed by multiplication of relative proportions to an indexing factor proportional to the microbial load of the samples. The module 'butyrate production II' describes butyrate production from the butyryl-CoA-acetate CoA-transferase pathway—the most common among colon bacteria. **a–d**, The abundance of the butyrate production II module was negatively correlated with BMI ($n = 771$ biologically independent samples, Spearman's $\rho = -0.27$, $P_{\text{adj}} = 3.1 \times 10^{-13}$) (**a**), fat mass percentage ($n = 771$ biologically independent samples, Spearman's $\rho = -0.21$, $P_{\text{adj}} = 6.0 \times 10^{-8}$) (**b**) and triglyceride levels ($n = 771$ biologically independent samples, Spearman's $\rho = -0.20$, $P_{\text{adj}} = 6.4 \times 10^{-8}$) (**c**), and significantly decreased in the Bact2 enterotype compared with the others (Bact2 < Prev < Bact1 = Rum; $n = 771$ biologically independent samples, Kruskal–Wallis $P_{\text{adj}} = 4.71 \times 10^{-35}$; different letters denote enterotypes with a significant pairwise difference (post hoc Dunn tests provided in Supplementary Table 10) (**d**). The body of the box plot represents the first and third quartiles of the distribution, the line represents the median, and the whiskers extend from the quartiles to the last data point within $1.5 \times \text{IQR}$, with outliers beyond. In **a–d**, adjustment for multiple testing (P_{adj}) was performed using the Benjamini–Hochberg method.



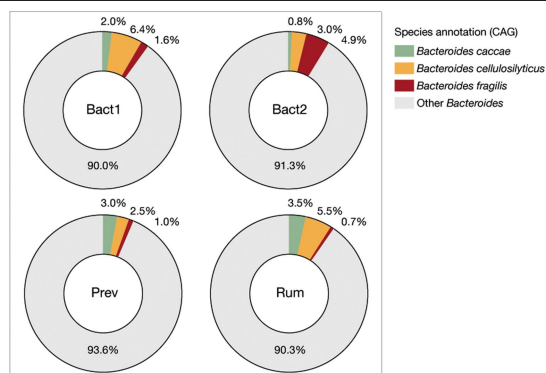
Extended Data Fig. 4 | Enterotyping of the MetaCardis dataset ($n = 2,022$ biologically independent samples). **a**, Principal coordinates visualization of the four enterotypes resulting from community typing was performed using DMM⁵² on genus-level faecal microbiome profiles. **b**, Information criteria (minimum Laplace) used to determine the optimal number of clusters (enterotypes) for the MetaCardis dataset ($n = 2,022$ biologically independent samples) DMM-based community typing. **c**, Average relative composition of

the enterotypes for key genera, used to label the enterotypes *Bacteroides*1 (Bact1; high percentages of *Bacteroides* and *Faecalibacterium*), *Bacteroides*2 (Bact2; high percentages of *Bacteroides* and low percentages of *Faecalibacterium*), *Prevotella* (Prev; high percentages of *Prevotella*) and *Ruminococcaceae* (Rum; low percentages of *Bacteroides* and *Prevotella*), on the basis of their respective genus-level proportional abundance profiles.

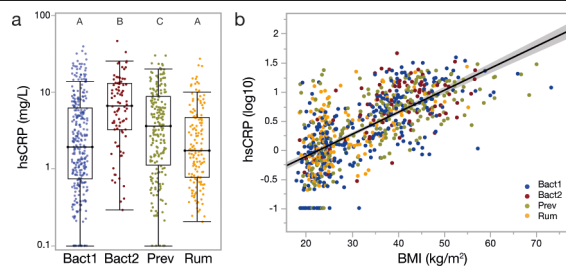


Extended Data Fig. 5 | Increased quantitative abundance of *Eggerthella* in the Bact2 enterotype of the non-statin-medicated BMIS cohort.

a, Difference in quantitative *Eggerthella* abundances between enterotypes (Prev = Rum < Bact1 < Bact2; $n = 771$ biologically independent samples, Kruskal-Wallis $P_{\text{adj}} = 4.10 \times 10^{-47}$; different letters denote enterotypes with a significant pairwise difference (post hoc Dunn tests provided in Supplementary Table 10)). Adjustment for multiple testing (P_{adj}) was performed using the Benjamini-Hochberg method. **b**, Difference in the proportion of *Eggerthella* (normalized by the sample total microbial load) between enterotypes, showing a comparable trend to that seen in **a** ($n = 771$ biologically independent samples). The body of the box plot represents the first and third quartiles of the distribution, the line represents the median, and the whiskers extend from the quartiles to the last data point within $1.5 \times \text{IQR}$, with outliers beyond.

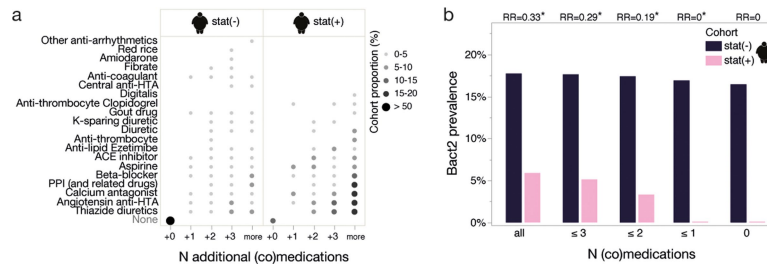


Extended Data Fig. 6 | Species dominating the *Bacteroides* fraction in the different enterotypes of the non-statin-medicated BMIS cohort. The top associations with the Bact2 enterotype—with the proportions they contribute to the total fraction shown in the ring chart—were the depletion in *B. caccae* ($n = 768$ biologically independent samples, Kruskal–Wallis, $P_{\text{adj}} = 1.3 \times 10^{-15}$) and *B. cellulosilyticus* ($n = 768$ biologically independent samples, Kruskal–Wallis, $P_{\text{adj}} = 5.3 \times 10^{-13}$) when compared with the Rum, Prev and Bact1 enterotypes, and the enrichment in *B. fragilis* ($n = 768$ biologically independent samples, Kruskal–Wallis, $P_{\text{adj}} = 3.5 \times 10^{-11}$; Supplementary Table 11). Species were defined by species-level annotation of metagenomic species, and their proportional abundances were defined relative to the genus abundance. Samples for which the genus had a low total abundance (below the 20% quantile for all species belonging to the top 10 genera) were excluded from the analysis ($n = 768$ biologically independent samples were included). Adjustment for multiple testing (P_{adj}) was performed using the Benjamini–Hochberg method.



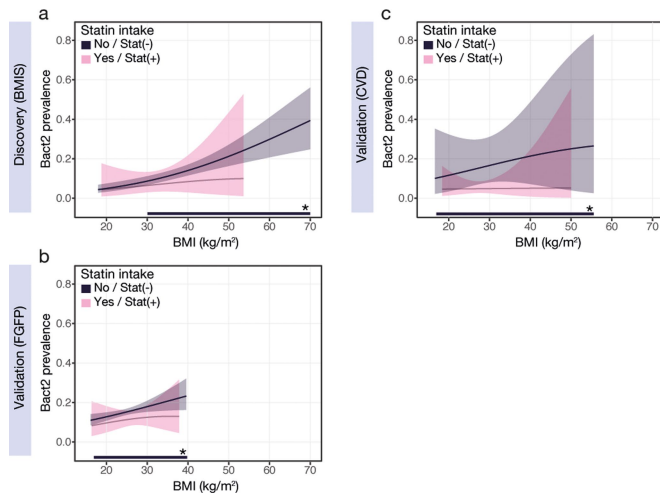
Extended Data Fig. 7 | Systemic inflammation and its relation to enterotypes and to BMI in the non-statin-medicated BMIS cohort.

a. Individuals with faecal samples enterotyped as Bact2 displayed more pronounced systemic inflammation levels as assessed through fasting serum hsCRP concentrations when compared with participants classified as Rum, Prev and Bact1 ($n = 763$ biologically independent samples, Kruskal–Wallis $P = 1.37 \times 10^{-10}$; Rum = Bact1 < Prev < Bact2; different letters denote enterotypes with a significant pairwise difference (post hoc Dunn tests provided in Supplementary Table 13)). The body of the box plot represents the first and third quartiles of the distribution, the line represents the median, and the whiskers extend from the quartiles to the last data point within $1.5 \times \text{IQR}$, with outliers beyond. **b.** Linear model of the correlation between host systemic inflammation (hsCRP concentration, \log_{10} -transformed) and BMI, fitted by least squares regression ($n = 763$ biologically independent samples; estimated intercept = -0.8681 , estimated slope = 0.0379 , $R^2 = 0.47$, $P = 1.5 \times 10^{-108}$).

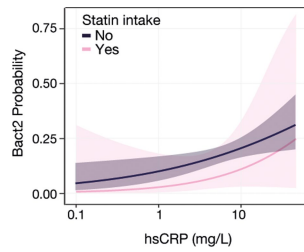


Extended Data Fig. 8 | Control for the effect of additional medication taken by obese statin-medicated or non-statin-medicated individuals of the BMIS cohort ($n = 888$ participants) on the association between reduced Bact2 prevalence and statin intake. **a, List of drugs taken by non-statin-medicated and statin-medicated obese BMIS participants separated into 5 groups: those reporting no (co-)medication (beyond statin intake) (+0), and those reporting one (+1), two (+2), three (+3) and more than three (more) (co-)medications. The size and colour of the dots represent the fraction of the non-statin-medicated or statin-medicated obese BMIS participants falling within that group. **b**, Difference in prevalence of the Bact2**

enterotype in statin-medicated compared with non-statin-medicated obese BMIS participants, with decreasing co-medication threshold for inclusion of participants. For 'all', the total number of statin-medicated and non-statin-medicated obese BMIS participants were included ($n = 474$ biologically independent samples); then only subjects reporting three or fewer (≤ 3 ; $n = 419$), two or fewer (≤ 2 ; $n = 369$), one or fewer (≤ 1 ; $n = 296$) or no (0; $n = 226$) (co-)medications were included. The relative risk and respective significance level associated with the prevalence of the Bact2 enterotype given statin intake is provided above the bar plots (Fisher's exact test, two-sided, $*P < 0.05$, relative risk = $P(\text{Bact2}|\text{Statin} = \text{Yes})/P(\text{Bact2}|\text{Statin} = \text{No})$).



Extended Data Fig. 9 | Variation in prevalence of the Bact2 enterotype with BMI and statin intake in the BMIS discovery cohort, and in the FGFP and CVD validation cohorts. a–c, Variation in the prevalence of the Bact2 enterotype with BMI for statin-medicated and non-statin-medicated individuals, showing the significant effect (represented by the range bar with an asterisk; Supplementary Table 16) of statin intake given individuals' BMI, in the BMIS obese participants ($n = 474$ biologically independent samples, multivariate binomial logistic regression, Statin | BMI, relative risk = 0.34, $*P_{\text{adj}} = 0.025$) (a); the FGFP cohort, a population-level recruitment with a much narrower BMI range than the BMIS cohort ($n = 2,345$ biologically independent samples, multivariate binomial logistic regression, Statin | BMI, relative risk = 0.72, $*P_{\text{adj}} = 0.045$) (b) and the MetaCardis CVD cohort ($n = 271$ biologically independent samples, excluding 11 individuals for which BMI was not known, multivariate binomial logistic regression, Statin | BMI, relative risk = 0.29, $*P_{\text{adj}} = 0.021$) (c). In a–c, the fit lines were obtained by multinomial logistic regression of enterotypes as predicted by BMI, for statin-medicated and non-statin-medicated individuals separately, with the shaded area corresponding to the 95% confidence intervals for the Bact2 regression. Adjustment for multiple testing (P_{adj}) was performed using the Benjamini–Hochberg method.



Extended Data Fig. 10 | Probability of carrying a Bact2 enterotype microbiota as a function of CRP levels and statin intake in the obese BMIS cohort. Association between systemic inflammation (measured by hsCRP levels) and having a faecal microbiota of the Bact2 enterotype, according to statin medication status. Binomial logistic regression (lines with 95% confidence intervals as shaded area) was performed for statin-medicated and non-statin-medicated individuals separately ($n = 462$ biologically independent samples).

Reporting Summary

Nature Research wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Research policies, see [Authors & Referees](#) and the [Editorial Policy Checklist](#).

Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

- | | |
|-------------------------------------|--|
| n/a | Confirmed |
| <input type="checkbox"/> | <input checked="" type="checkbox"/> The exact sample size (n) for each experimental group/condition, given as a discrete number and unit of measurement |
| <input type="checkbox"/> | <input checked="" type="checkbox"/> A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly |
| <input type="checkbox"/> | <input checked="" type="checkbox"/> The statistical test(s) used AND whether they are one- or two-sided
<i>Only common tests should be described solely by name; describe more complex techniques in the Methods section.</i> |
| <input type="checkbox"/> | <input checked="" type="checkbox"/> A description of all covariates tested |
| <input type="checkbox"/> | <input checked="" type="checkbox"/> A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons |
| <input type="checkbox"/> | <input checked="" type="checkbox"/> A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals) |
| <input type="checkbox"/> | <input checked="" type="checkbox"/> For null hypothesis testing, the test statistic (e.g. F , t , r) with confidence intervals, effect sizes, degrees of freedom and P value noted
<i>Give P values as exact values whenever suitable.</i> |
| <input checked="" type="checkbox"/> | <input type="checkbox"/> For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings |
| <input type="checkbox"/> | <input checked="" type="checkbox"/> For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes |
| <input type="checkbox"/> | <input checked="" type="checkbox"/> Estimates of effect sizes (e.g. Cohen's d , Pearson's r), indicating how they were calculated |

Our web collection on [statistics for biologists](#) contains articles on many of the points above.

Software and code

Policy information about [availability of computer code](#)

Data collection	[Metagenomic data] Sequencing reads, obtained from ion-proton technology (ThermoFisher Scientific), were cleaned using Alien Trimmer (v.0.4.0). Filtered reads were mapped to the 9.9 million-gene catalogue using Bowtie (v2.2.6). A gene abundance table was generated by means of a two-step procedure using METEOR, and posterior normalization was performed using the MetaOMineR R package using the FPKM strategy. Metagenomic species (MGS) or co-abundant gene groups were previously computed on 1267 human gut metagenomes used to construct the 9.9 million-gene catalogue, their abundance profiles for the MetaCardis cohort was estimated as the mean abundance of the 50 genes defining a robust centroid of the cluster, with taxonomic assignment by NCBI blast N on the NCBI database (November 2016 version). [Faecal microbial loads] Flow cytometry analysis was performed using a C6 Accuri flow cytometer (BD Biosciences, New Jersey, USA), using the BD Accuri CFlow software (v1.0.264.21) for gating and events counting.
Data analysis	[Metagenomic data] Gut metabolic modules (GMM) profiles were calculated using the software Omixer-RPM v1.0 and the newest version of the GMMs (v.2.0), which includes a specific set of six modules zooming in on bacterial TMA metabolism. [Quantitative microbiota profiles] QMP profiles were created using the phyloseq R package to rarefy the profiles to even sampling depth, sampling depth, defined as the ratio between sampling size (average mOTU marker genes coverage ⁴⁴) and microbial load (average total cell count per gram of frozen faecal material). [Statistical analyses] Statistical analyses were performed on Rstudio v1.1.456, using the following R packages: vegan v2.5-3, phyloseq v1.26.0, FSA v0.8.24, coin v1.2-2, DirichletMultinomial v1.24.0, Hmisc v4.1-1, car v3.0-2, sjstats v0.17.5, and nnet v7.3-12.

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors/reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Research [guidelines for submitting code & software](#) for further information.

Data

Policy information about [availability of data](#)

All manuscripts must include a [data availability statement](#). This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A list of figures that have associated raw data
- A description of any restrictions on data availability

Raw amplicon sequencing data used in this study have been deposited in the EMBL-EBI European Nucleotide Archive (ENA) under accession number PRJEB37249 [public access]. The metadata and processed microbiome data required for re-analysis of results presented in the manuscript are respectively provided as Extended Data 1 (tab separated file) and downloadable at <http://raeslab.org/software/BMIS/>.

Field-specific reporting

Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

☒ Life sciences ☐ Behavioural & social sciences ☐ Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see nature.com/documents/nr-reporting-summary-flat.pdf

Life sciences study design

All studies must disclose on these points even when the disclosure is negative.

Sample size	No sample size calculation was performed prior to cohort recruitment. Based on the baseline prevalence of Bact2 enterotype (with baseline defined as lean/overweight individuals P(Bact2)=14%) in the amplicon-sequenced Flemish Gut Flora Project cohort, the present study cohort size allowed to identify a minimum difference of 7.4% in Bact2 prevalence between the two groups lean/overweight (N=414) vs obese (N=474) as significant (power=80%, alpha=0.05). Because the FGFP cohort was a population microbiome monitoring effort, while the BMIS cohort was actively recruited to have a balanced representation over a wide BMI range, therefore the sample size would be sufficient to detect a smaller prevalence difference.
Data exclusions	This study targeted the analysis of metabolic alterations associated with body mass index ranging from normal to severe obesity. The BMIS cohort (N=888) was selected from the MetaCardis consortium study cohort, by exclusion of cardiovascular patients (as defined in the MetaCardis consortium study protocol as patient groups 4, 5, 6 and 7) and any individual with type-2 diabetes (T2D). T2D diagnosis was defined using the American Diabetes Association (ADA) definition: fasting glycemia > 6.9 mmol/l and/or 2h values in the oral glucose tolerance test > 11 mmol/l and/or haemoglobin A1c (HbA1c, glycated haemoglobin) ≥ 6.5% and/or use of any anti-diabetic treatment. but excluding volunteers with a diagnosis of diabetes as to avoid the potentially confounding effect of the associated medication. This exclusion criteria was pre-established for this manuscript. The same criteria for inclusion and exclusion that were used for MetaCardis recruitment were applied to the FGFP validation cohort: inclusion of age ranging from 18 to 75 years old, exclusion of acute or chronic inflammatory or infectious diseases – notably diagnosis of inflammatory bowel disease and recent gastro-enteritis, and exclusion of diabetes patients – defined as above as diabetes diagnosis or elevated glycated haemoglobin (HbA1c ≥ 6.5%) or use of any anti-diabetic treatment.
Replication	All microbiome observations regarding Bact2 prevalence associated with obesity and statin therapy were successfully replicated in another subset of the MetaCardis cohort - the cardiovascular patients [CVD] (as defined in the MetaCardis consortium study protocol as patient groups 4, 5, 6 and 7), and in an independent cross-sectional cohort - the Flemish Gut Flora Project cohort [FGFP].
Randomization	Not applicable: this was a cross-sectional study, not a randomized study. No intervention was performed on subjects, and therefore no random allocation into groups. Potentially confounding covariates were identified as variables with significant association to the response or dependent variables and were added to a multivariate model to validate the findings.
Blinding	Not applicable: this was a cross-sectional study, not a randomized study. The investigators were not blinded during data collection or data analysis.

Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

Materials & experimental systems

n/a	Involved in the study
<input checked="" type="checkbox"/>	<input type="checkbox"/> Antibodies
<input checked="" type="checkbox"/>	<input type="checkbox"/> Eukaryotic cell lines
<input checked="" type="checkbox"/>	<input type="checkbox"/> Palaeontology
<input checked="" type="checkbox"/>	<input type="checkbox"/> Animals and other organisms
<input type="checkbox"/>	<input checked="" type="checkbox"/> Human research participants
<input type="checkbox"/>	<input checked="" type="checkbox"/> Clinical data

Methods

n/a	Involved in the study
<input checked="" type="checkbox"/>	<input type="checkbox"/> ChIP-seq
<input type="checkbox"/>	<input checked="" type="checkbox"/> Flow cytometry
<input checked="" type="checkbox"/>	<input type="checkbox"/> MRI-based neuroimaging

Human research participants

Policy information about [studies involving human research participants](#)

Population characteristics	A complete description of the study participants can be found in Supplementary Table 1 (blood panel and current medication intake). The BMIS cohort consisted of N=888 participants (574 Females, 314 Males), with a median BMI of 31.49 [17.95-73.26] and a median age of 54 [18-76], recruited in 3 countries (294 DE, 247 DK, 347 FR). None of the included participants were diagnosed as diabetic.
Recruitment	The N=888 transnational Body Mass Index Spectrum (BMIS) cohort was assembled as part of the overall MetaCardis recruitment efforts. Participants were recruited between 2013 and 2015 in the clinical departments of the Pitié-Salpêtrière Hospital (Paris, France), the Integrated Research and Treatment Center for Adiposity Diseases (Leipzig, Germany), and in the Novo Nordisk Foundation Center for Basic Metabolic Research (Copenhagen, Denmark). Potential participants were evaluated for suitability according to standardized inclusion and exclusion criteria across the three recruitment centers. Exclusion criteria included history of abdominal cancer/radiation therapy on the abdomen, history of intestinal resection (except for appendectomy), acute or chronic inflammatory or infectious diseases (including VHC, VHB, and HIV), history of organ transplantation or receiving immunosuppressive therapy, severe kidney failure (MDRD glomerular filtration rate <50 ml (min 1.73m ²)-1), or drug or alcohol addiction. All study participants had to be free of any antibiotic use in the three months prior to inclusion. We do not expect any (self-)selection bias that would have an impact on the results.
Ethics oversight	Ethical approval was obtained from the Ethics Committee CPP Ile-de France, Ethics Committee at the Medical Faculty of the University of Leipzig, and the Ethical Committees of the Capital Region of Denmark. Study design complied with all relevant ethical regulations, aligning with the Helsinki Declaration and in accordance with European privacy legislation. All participants provided written informed consent.

Note that full information on the approval of the study protocol must also be provided in the manuscript.

Clinical data

Policy information about [clinical studies](#)

All manuscripts should comply with the ICMJE [guidelines for publication of clinical research](#) and a completed [CONSORT checklist](#) must be included with all submissions.

Clinical trial registration	The study protocol was registered at clinicaltrial.gov (NCT02059538).
Study protocol	The study protocol is available at https://clinicaltrials.gov/ct2/show/NCT02059538
Data collection	The N=888 transnational Body Mass Index Spectrum (BMIS) cohort was assembled as part of the overall MetaCardis recruitment efforts. Participants were recruited between 2013 and 2015 in the clinical departments of the Pitié-Salpêtrière Hospital (Paris, France), the Integrated Research and Treatment Center for Adiposity Diseases (Leipzig, Germany), and in the Novo Nordisk Foundation Center for Basic Metabolic Research (Copenhagen, Denmark).
Outcomes	The hypotheses tested in this manuscript were not listed as part of the planned NCT02059538 study outcomes. The primary predefined outcome of the MetaCardis project (description of differences in gut microbiota signatures between MetaCardis study groups using metagenomic sequencing) is not addressed in the present manuscript.

Flow Cytometry

Plots

Confirm that:

- ☒ The axis labels state the marker and fluorochrome used (e.g. CD4-FITC).
- ☒ The axis scales are clearly visible. Include numbers along axes only for bottom left plot of group (a 'group' is an analysis of identical markers).
- ☒ All plots are contour plots with outliers or pseudocolor plots.
- ☒ A numerical value for number of cells or percentage (with statistics) is provided.

Methodology

Sample preparation	0.2 g frozen (-80°C) faecal aliquots were dissolved in physiological solution to a total volume of 100 mL (8.5 g/L NaCl). Subsequently, the slurry was diluted 1,000 times. Samples were filtered using a sterile syringe filter (pore size of 5 µm). 1 mL of the microbial cell suspension obtained was stained with 1 µL SYBR Green I (1:100 dilution in DMSO; shaded 15 min incubation at 37°C).
Instrument	C6 Accuri flow cytometer (BD Biosciences, New Jersey, USA).
Software	BD Accuri CFlow software v1.0.264.21 (BD Biosciences, New Jersey, USA).
Cell population abundance	Not applicable. No sorting of fractions was performed.
Gating strategy	Fluorescence events were monitored using the FL1 533/30 nm and FL3 >670 nm optical detectors. In addition, also forward and sideward-scattered light was collected. The BD Accuri CFlow software was used to gate and separate the microbial fluorescence events on the FL1/FL3 density plot from background. A threshold value of 2000 was applied on the FL1 channel. The gated fluorescence events were evaluated on the forward/sideward density plot, as to exclude remaining background events. Instrument and gating settings were kept identical for all samples.

☒ Tick this box to confirm that a figure exemplifying the gating strategy is provided in the Supplementary Information.

TASL is the SLC15A4-associated adaptor for IRF5 activation by TLR7–9

<https://doi.org/10.1038/s41586-020-2282-0>

Received: 3 October 2019

Accepted: 7 April 2020

Published online: 13 May 2020

 Check for updates

Leonhard X. Heinz¹, JangEun Lee², Utkarsh Kapoor¹, Felix Kartnig¹, Vitaly Sedlyarov¹, Konstantinos Papakostas¹, Adrian César-Razquin¹, Patrick Essletzbichler¹, Ulrich Goldmann¹, Adrijana Stefanovic¹, Johannes W. Bigenzahn¹, Stefania Scorzoni¹, Mattia D. Pizzagalli¹, Ariel Bensimon¹, André C. Müller¹, F. James King², Jun Li², Enrico Girardi¹, M. Lamine Mbow², Charles E. Whitehurst², Manuele Rebsamen^{1✉} & Giulio Superti-Furga^{1,3✉}

Toll-like receptors (TLRs) have a crucial role in the recognition of pathogens and initiation of immune responses^{1–3}. Here we show that a previously uncharacterized protein encoded by *CXorf21*—a gene that is associated with systemic lupus erythematosus^{4,5}—interacts with the endolysosomal transporter SLC15A4, an essential but poorly understood component of the endolysosomal TLR machinery also linked to autoimmune disease^{4,6–9}. Loss of this type-I-interferon-inducible protein, which we refer to as ‘TLR adaptor interacting with SLC15A4 on the lysosome’ (TASL), abrogated responses to endolysosomal TLR agonists in both primary and transformed human immune cells. Deletion of SLC15A4 or TASL specifically impaired the activation of the IRF pathway without affecting NF- κ B and MAPK signalling, which indicates that ligand recognition and TLR engagement in the endolysosome occurred normally. Extensive mutagenesis of TASL demonstrated that its localization and function relies on the interaction with SLC15A4. TASL contains a conserved pLxIS motif (in which p denotes a hydrophilic residue and x denotes any residue) that mediates the recruitment and activation of IRF5. This finding shows that TASL is an innate immune adaptor for TLR7, TLR8 and TLR9 signalling, revealing a clear mechanistic analogy with the IRF3 adaptors STING, MAVS and TRIF^{10,11}. The identification of TASL as the component that links endolysosomal TLRs to the IRF5 transcription factor via SLC15A4 provides a mechanistic explanation for the involvement of these proteins in systemic lupus erythematosus^{12–14}.

Eukaryotic cells recognize a large variety of pathogens using a limited set of receptors, adaptors and signalling molecules in a modular and combinatorial fashion, integrating information on subcellular localization and metabolism, to trigger appropriate responses^{15–17}. While endolysosomal TLRs have evolved to respond to microbial nucleic acids of various origins, aberrant activation can contribute to development of autoimmunity². Indeed, many TLR pathway components are genetically linked to a predisposition to autoimmune diseases, such as systemic lupus erythematosus (SLE)^{12–14}. We studied SLC15A4, a member of the solute carrier family that has previously been implicated in endolysosomal TLR activation and autoimmune diseases through mouse models and human genetics^{4,6–9,18–23}. However, the mechanistic role and hierarchical position of SLC15A4 in the TLR pathway has not yet been fully elucidated. The human THP1 monocytic cell line expresses both lysosomal members of the SLC15 family, SLC15A3 and SLC15A4 (Extended Data Fig. 1a). CRISPR–Cas9-mediated inactivation of SLC15A4 abrogated TNF production upon stimulation with R848 (a specific agonist of TLR7 and TLR8), demonstrating a non-redundant role for SLC15A4 in human monocytic cells (Fig. 1a). To gain a mechanistic understanding of how SLC15A4 affects this process,

we set out to identify its binding partners by interaction proteomics using tandem-affinity purification (TAP) coupled to gel-free liquid chromatography–tandem mass spectrometry (LC–MS/MS)^{24,25}. We generated THP1 cells stably expressing a tagged lysosomal SLC15A4 wild-type glycoprotein, a construct that lacks a large cytosolic loop (deletion of amino acids 253–303, SLC15A4(Δ loop)) or a construct in which the cytosolic N terminus is deleted (deletion of amino acids 1–28, SLC15A4(Δ N)). SLC15A4(Δ N) lacks a previously described di-leucine-containing motif (L14–L15) that mediates lysosomal sorting and is therefore mislocalized to the plasma membrane^{7,26} (Extended Data Fig. 1b–e). TAP–MS/MS analysis of these variants provided a global view of the SLC15A4 interaction landscape, and revealed CXorf21—a previously uncharacterized 301-amino-acid protein conserved in vertebrates—as a prominent and specific binder (Fig. 1b, Extended Data Fig. 2a, Supplementary Table 1). *CXorf21* is an X-chromosome-encoded gene, which has previously been shown to be genetically linked to SLE and hypothesized to contribute to the sexual dimorphism of this disease^{4,5,27,28}. On the basis of evidence presented here, we refer to the protein encoded by the *CXorf21* gene as ‘TLR adaptor interacting with SLC15A4 on the lysosome’ (TASL). While SLC15A4 shows a wider tissue

¹CeMM Research Center for Molecular Medicine of the Austrian Academy of Sciences, Vienna, Austria. ²Boehringer Ingelheim Pharmaceuticals, Ridgefield, CT, USA. ³Center for Physiology and Pharmacology, Medical University of Vienna, Vienna, Austria. ✉e-mail: mrebsamen@cemm.oew.ac.at; gsuperti@cemm.oew.ac.at

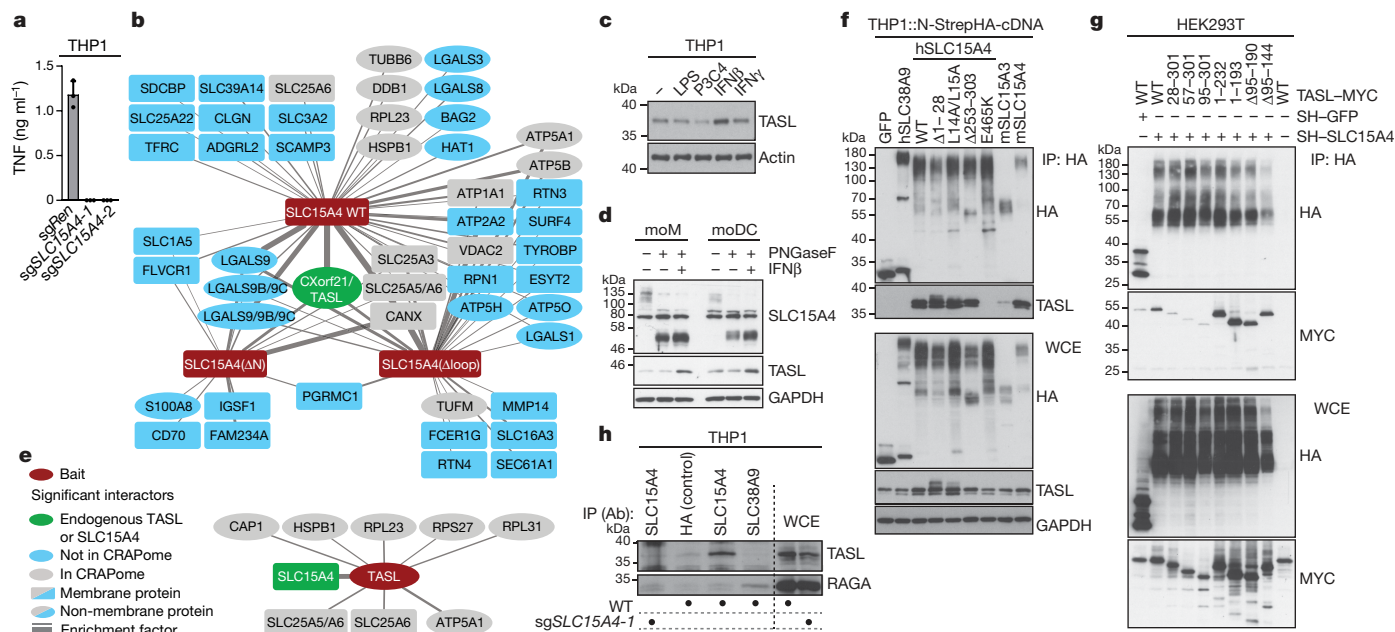


Fig. 1 | TASP, a protein inducible by type I interferons, is a specific interaction partner of SLC15A4. **a**, TNF production of indicated THP1 cells stimulated with R848 (5 $\mu\text{g ml}^{-1}$, for 24 h). Mean \pm s.d. ($n = 3$ biological replicates). sgRen, single guide (sg)RNA targeting *Renilla*; sgSLC15A4-1 and sgSLC15A4-2, two different sgRNAs targeting *SLC15A4*. **b**, **e**, Interaction networks of SLC15A4 and deletion mutants (**b**) and TASP (**e**) identified by TAP-MS/MS. Baits are shown in red. Prey proteins (false discovery rate (FDR) calculated using 'Significance Analysis of Interactome' (SAINT) of $<1\%$) are shown in blue or grey if present in the 'Contaminant Repository for Affinity Purification' (CRAPome) database. Interactions are represented as edges, and line width corresponds to the enrichment factor calculated by SAINT. WT, wild type. **c**, Immunoblots of THP1 cells stimulated (for 16 h) with

lipopolysaccharide (LPS) (100 ng ml^{-1}), Pam3CSK4 (P3C4) (100 ng ml^{-1}), interferon- β (20 ng ml^{-1}) or interferon- γ (20 ng ml^{-1}). **d**, Immunoblots of lysates from monocyte-derived macrophages (moM) and dendritic cells (moDC) stimulated with interferon- β (20 ng ml^{-1} , for 16 h) treated with PNGase F, as indicated. **f**, **g**, Immunoprecipitates (IP, haemagglutinin tag (HA)) and whole-cell extracts (WCE) from transduced THP1 (**f**) or transiently transfected HEK293T (**g**) cells analysed by immunoblotting. h, human; m, mouse; SH, Strep-HA tag. **h**, Immunoprecipitates (indicated antibodies) and whole-cell extracts from indicated THP1 cells were analysed by immunoblotting. In **a**, **c**, **d**, **f**–**h**, data are representative of five (**a**) or two (**c**, **d**, **f**–**h**) independent experiments. For gel source data, see Supplementary Fig. 1.

distribution, TASP appears to be restricted to the haematopoietic compartment and, in particular, to myeloid cells, B lymphocytes and plasmacytoid dendritic cells^{29–31} (Extended Data Fig. 2b, c, Supplementary Table 2). Furthermore, expression of TASP and—to a lesser extent—SLC15A4 was induced by treatment with interferon- β (Fig. 1c, Extended Data Fig. 3a, b). Considering the relevance of type I interferons in SLE^{12,14,32}, we further confirmed this observation in primary human monocyte-derived macrophages and dendritic cells (Fig. 1d). Together, these data suggested that TASP could be specifically involved in the immune-related functions of SLC15A4.

Next, we characterized the SLC15A4–TASP protein complex. We performed TAP-MS/MS analysis using TASP as bait, which identified endogenous SLC15A4 (Fig. 1e). Conversely, tagged SLC15A4 immunoprecipitated endogenous TASP in two different cellular systems (Fig. 1f, Extended Data Fig. 3c). SLC15A4 binding required the N-terminal region of TASP (Fig. 1g, Extended Data Fig. 3d). Demonstrating endogenous complex formation and specificity, we detected TASP in SLC15A4 immunoprecipitates from wild-type, but not SLC15A4-deficient, cells. Moreover, we did not observe TASP upon immunoprecipitation of lysosomal SLC38A9 (which recovered its binding partner RAGA²⁴) (Fig. 1h, Extended Data Fig. 1d). SLC15A4, but not the closely related SLC15A3, interacted with TASP (Fig. 1f, Extended Data Figs. 1d, 3e, h). Mutant forms of SLC15A4 (SLC15A4(ΔN) and SLC15A4(L14A/L15A)) that mislocalize to the plasma membrane retained binding and led to the accumulation of a phosphatase-sensitive, slower-migrating form of TASP, indicating that the interaction was independent of the subcellular context (Fig. 1f, Extended Data Figs. 1d, 3f). By contrast, a point mutation

(E465K) in SLC15A4 that affects a conserved glutamate residue (which has previously been shown to be required for substrate binding and transport⁹) resulted in complete loss of TASP binding, raising the possibility that the interaction is conformation-dependent (Fig. 1f, Extended Data Fig. 1b, d, e).

The expression of SLC15A4 constructs that are able to bind TASP resulted in an increase in the abundance of TASP, whereas SLC15A4-knockout cells showed reduced levels of endogenous TASP (Figs. 1f, 2a, Extended Data Figs. 3c, 5e). Furthermore, co-expression of wild-type SLC15A4 or SLC15A4(ΔN) in THP1 cells that stably express TASP tagged with green fluorescent protein (TASP–GFP) led to a strong increase in GFP signal, and to the recruitment of TASP–GFP to endolysosomal structures or the plasma membrane, respectively (Extended Data Fig. 4a–d). By contrast, co-expression of SLC15A4(E465K) only marginally affected TASP–GFP levels or localization. Together, these experiments revealed a proteostatic relationship that regulates TASP abundance depending on SLC15A4 expression levels and binding.

We then assessed the relevance of the SLC15A4–TASP module for TLR-induced inflammatory responses. While no major effects were observed on steady-state gene expression, SLC15A4 or TASP deficiency blunted transcriptional responses to R848 stimulation (Fig. 2a–d, Extended Data Fig. 5a–c, Supplementary Table 3). Mirroring SLC15A4 deletion, TASP-knockout cells showed a strong impairment in R848-induced production of cytokines and chemokines, upregulation of PD-L1 (also known as CD274) and activation of signalling pathway-specific reporters (Fig. 2e, f, Extended Data Fig. 5d–f). Similar defects were observed using other TLR8 ligands, whereas knockout

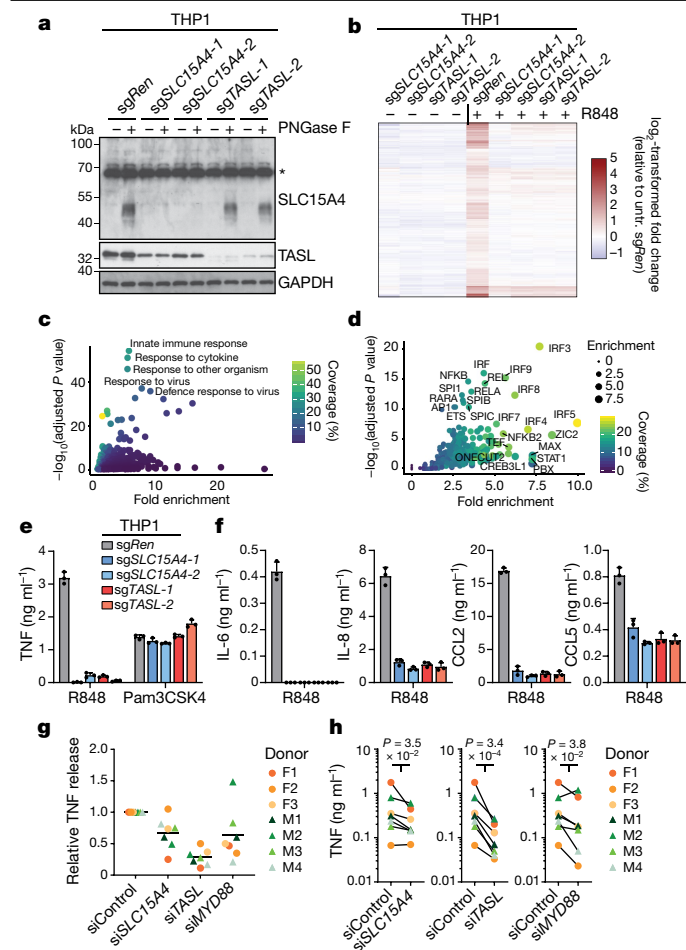


Fig. 2 | TASL and SLC15A4 are required for the function of endolysosomal TLR7 and TLR8. **a**, Immunoblots of THP1 cell lines. Lysates treated with PNGase F as indicated. Asterisk denotes a non-specific band. **sgTASL-1** and **sgTASL-2**, two different sgRNAs targeting *CXorf21*. **b**, Transcriptional profiles of untreated (untr.) and R848-treated ($5 \mu\text{g ml}^{-1}$, for 6 h) THP1 cell lines. Genes significantly upregulated (DESeq2 adjusted P value < 0.05 , $n = 3$ biological replicates) upon treatment with R848 in control (sgRen) are shown. **c**, Gene Ontology (GO) enrichment analysis (two-sided Fisher's exact test, P value adjusted for multiple testing) for R848-induced genes in control THP1 cells as defined in **b**. x-axis, fold enrichment of GO terms in the set of upregulated genes compared to all genes expressed (counts per million > 1). y-axis, significance of enrichment ($-\log_{10}$ -transformed P value adjusted for multiple testing). Colour denotes the fraction of R848-induced genes included in the corresponding GO term. **d**, Transcription factor enrichment analysis (two-sided Fisher's exact test, P value adjusted for multiple testing) of R848-induced genes in control THP1 cells as defined in **b**. Background set, all expressed genes (counts per million > 1). **e**, **f**, Cytokine production of THP1 cells stimulated (for 24 h) with R848 ($5 \mu\text{g ml}^{-1}$) or Pam3CSK4 (100 ng ml^{-1}). Mean \pm s.d. ($n = 3$ biological replicates). **g**, **h**, TNF production of CD14⁺ monocytes transfected with short interfering (si)RNA against *SLC15A4* (siSLC15A4), *CXorf21* (siTASL) or *MYD88* (siMYD88), and stimulated with R848 ($5 \mu\text{g ml}^{-1}$, for 24 h) from seven healthy donors. F, female; M, male. Circles represent relative (normalized to control siRNA (siControl)) (g) or absolute (h) TNF levels from seven healthy donors as mean of triplicates. In g, lines indicate mean over seven donors. **h**, Difference in \log_2 -transformed TNF concentrations, relative to control siRNA, of all seven donors tested using paired, two-sided t -test. In **a**, **e**, **f**, data are representative of two independent experiments. For gel source data, see Supplementary Fig. 1.

cells responded normally to agonists of plasma-membrane-localized TLR2 and TLR5, as well as agonists of the STING-dependent cytoplasmic DNA-sensing pathway, which demonstrates the specificity of TASL–SLC15A4 for endolysosomal TLRs (Fig. 2e, Extended Data Fig. 5f, g).

In contrast to SLC15A4, knockout of the related SLC15A2 and SLC15A3 did not impair R848-induced pathway activation (Extended Data Fig. 5h). The response to type I interferons was intact in cells deficient in TASL or SLC15A4, which indicates that impaired interferon signalling is not the underlying cause of the observed phenotypes (Extended Data Fig. 5i). Additionally, no defects in NOD1 or NOD2 responses were detected in the absence of TASL or SLC15A4, possibly because other transporters act redundantly to mediate ligand uptake^{8,26,33,34} (Extended Data Fig. 5f, j, k). To further define the relevance of the TASL–SLC15A4 module, we investigated primary human cells. Knockdown of SLC15A4 and—even more prominently—TASL resulted in the reduction of TNF production upon R848 stimulation of primary CD14⁺ monocytes, highlighting the role of the complex for endolysosomal TLR responses (Fig. 2g, h, Extended Data Fig. 5l).

Similar to TLR7 and TLR8, TLR9 (the other endolysosomal TLR linked to SLC15A4⁷) has a central role in physiological and pathological immune responses². Although THP1 cells respond poorly to TLR9 agonists, stable expression of the receptor resulted in secretion of pro-inflammatory mediators and interferon- β upon stimulation with CpG-A or CpG-B (Extended Data Fig. 6a, b). Deletion of SLC15A4 or TASL markedly impaired all these responses (Fig. 3a, Extended Data Fig. 6a, c). Cells deficient in SLC15A4 or TASL did not display overt defects in levels or processing of TLR proteins, lysosomal protein abundance nor any detectable alterations in ligand uptake or lysosomal acidification (Extended Data Fig. 6a, d–k). We then monitored endolysosomal TLR-induced signalling. STAT1 activation, which is probably induced by paracrine interferon, was strongly diminished in SLC15A4- and TASL-knockout cells upon treatment with CpG or R848 (Fig. 3b, Extended Data Fig. 6l). No major defects in the activation of the NF- κ B and MAPK pathways were detected (Fig. 3b, Extended Data Fig. 6l). This result indicates that TLR engagement occurred normally in cells deficient in TASL or SLC15A4, and therefore places the complex downstream of early ligand–receptor activation events. Altogether these data suggested a defect in the activation of the IRF pathway and pointed to a possible involvement of IRF5, given its role in the induction of proinflammatory mediators and type I interferons downstream of TLR7, TLR8 and TLR9 as well as its clear association with SLE^{35–39}. Indeed, knockout of IRF5—but not of IRF3 or IRF7—specifically compromised responses to endolysosomal TLR agonists in THP1 cells (Extended Data Fig. 7a–g). Accordingly, we found that IRF5 deficiency had the strongest effect on CpG-induced gene expression, when compared to IRF3 or IRF7 (Fig. 3c, Extended Data Fig. 7h, Supplementary Table 4). We assessed IRF5 activation upon CpG stimulation and observed that loss of SLC15A4 or TASL impaired IRF5 phosphorylation (Fig. 3b). Importantly, defects in the CpG-induced transcriptional responses observed in TASL- or SLC15A4-knockout cells mirrored those of IRF5-deficient cells, which strongly supports an epistatic relationship between the function of the SLC15A4–TASL complex and IRF5 (Fig. 3d–f, Extended Data Fig. 7h–j). Indeed, genes affected by a deficiency of TASL or SLC15A4 were enriched in IRF targets whereas this was not the case in the subset of unaffected transcripts, which displayed a signature related to NF- κ B (Fig. 3f, Extended Data Fig. 7j, k).

Plasmacytoid dendritic cells are major producers of type I interferons, respond efficiently to endosomal TLR activation and contribute to the pathogenesis of SLE^{13,32}. In light of this and the previously described role of SLC15A4 in these cells^{7,18,19}, we investigated the function of the SLC15A4–TASL complex in the human plasmacytoid dendritic cell line CAL-1^{35,40,41}. Knockout cells displayed impaired cytokine production upon activation of endogenous TLR7, TLR8 and TLR9, which was associated with a specific defect in the activation of IRF5 but not of NF- κ B or MAPK pathways (consistent with the phenotypes observed in THP1 monocytes) (Fig. 3g–i, Extended Data Fig. 8a–g). To define the molecular mechanism by which the TASL–SLC15A4 complex controls IRF5-dependent endolysosomal TLR

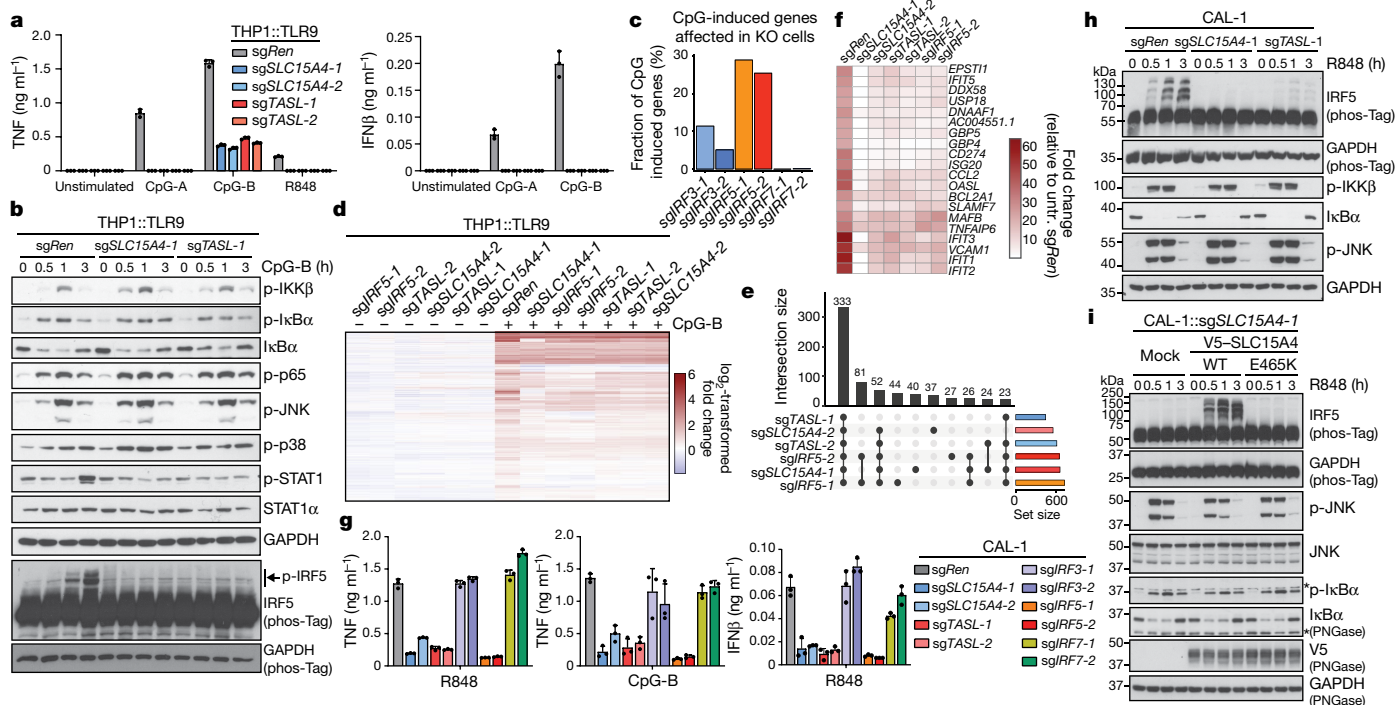


Fig. 3 | TASL and SLC15A4 deficiency selectively impairs IRF5-dependent endolysosomal TLR signalling. **a**, Cytokine production of indicated TLR9-expressing THP1 cells (THP1::TLR9) unstimulated or stimulated (for 24 h) with CpG-A (5 μ M), CpG-B (5 μ M) or R848 (5 μ g ml⁻¹). **b**, Immunoblots of indicated THP1::TLR9 cells stimulated with CpG-B (5 μ M, for 0–3 h). p-, phosphorylated. **c**, Fraction of genes induced by CpG-B (2 μ M, for 6 h) (DESeq2 adjusted *P* value < 0.05, *n* = 3 biological replicates) affected in *IRF3*, *IRF5* or *IRF7*-deficient THP1::TLR9 cells compared to control (sgRen). sgIRF3-1 and sgIRF3-2, two different sgRNAs targeting *IRF3*; sgIRF5-1 and sgIRF5-2, two different sgRNAs targeting *IRF5*; sgIRF7-1 and sgIRF7-2, two different sgRNAs targeting *IRF7*. KO, knockout. **d**, Transcriptional profiles of unstimulated and CpG-B-treated (2 μ M, for 6 h) THP1::TLR9 cell lines. Genes significantly

upregulated (DESeq2 adjusted *P* value < 0.05, *n* = 3 biological replicates) by CpG-B in control (sgRen) are shown. **e**, Upset plot representing the number of CpG-B-induced genes that are commonly affected by sgRNAs; the ten largest sets are shown. **f**, Heat map representing 20 most-induced genes by CpG-B in control THP1::TLR9 cells that are significantly (DESeq2 adjusted *P* value < 0.05, *n* = 3 biological replicates) affected by *SLC15A4*, *CXorf21* and *IRF5* knockout (related to **d**, **e**). **g**, Cytokine production of indicated CAL-1 cells stimulated (for 24 h) with R848 (5 μ g ml⁻¹) or CpG-B (5 μ M). **h**, **i**, Immunoblots of knockout (**h**) or reconstituted (**i**) CAL-1 cells stimulated with R848 (5 μ g ml⁻¹, for 0–3 h), as indicated. In **a**, **g**, mean \pm s.d. (*n* = 3 biological replicates). In **a**, **g**–**i**, data are representative of two independent experiments. For gel source data, see Supplementary Fig. 1.

signalling, we reconstituted knockout cells. Expression of wild-type SLC15A4, but not SLC15A3, rescued the impaired R848 responses as well as the diminished levels of TASL (Fig. 4a, Extended Data Fig. 8h, i). This activity required the SLC15A4 transmembrane core (Extended Data Fig. 8h, i). SLC15A4 variants that localize to the plasma membrane rescued the abundance of TASL protein, but not signalling (Fig. 4a). In addition to correct localization, TASL-binding was required to restore endolysosomal TLR responses, as shown by the fact that the SLC15A4(E465K) mutant did not rescue knockout cells (Figs. 3i, 4a). Furthermore, substitution of E465 with alanine confirmed the critical involvement of this residue in both TASL-binding and function (Extended Data Fig. 8j). By contrast, mutation of glutamate residues (E44 or E47) in the conserved ExxER motif, which has previously been shown to be required for proton-coupled transport in multiple transporters related to SLC15A4⁴², retained TASL-binding and functional rescuing capabilities (Extended Data Fig. 8j). These data suggest a requirement for the TASL-binding, but not necessarily the transport activity, of SLC15A4 for endosomal TLR function. We next profiled the entire TASL protein sequence using a series of 50 sequential mutants (which we numbered as mutants 1–50), in which polar residues were exchanged with alanine and stably expressed in knockout cells (Extended Data Fig. 9a–d). We identified several evolutionarily conserved elements that were required for endolysosomal TLR-induced responses (Fig. 4b). Whereas the N-terminal mutant 1 (covering amino acids 1–8) did not bind to endogenous SLC15A4, the interaction was retained by

the defective mutants in which the central (amino acids 106–128) and C-terminal (amino acids 204–296) regions were targeted (Fig. 4b, Extended Data Fig. 9e). These data suggest a role for the central and C-terminal elements in mediating the effector functions of TASL. Accordingly, TASL functionality was compromised by N- or C-terminal tagging (Extended Data Fig. 9f). Sequence analysis of the C-terminal region of TASL revealed homology with the IRF5 C-terminal domain, which is required for phosphorylation-induced activation and dimerization^{43,44} (Fig. 4c, d). Upon inspection of the homologous region of TASL, we identified a highly conserved pLxIS motif, which mediates in the canonical innate adaptors STING, MAVS and TRIF the phosphorylation-dependent homotypic recruitment and activation of IRF3 downstream of their respective pattern recognition receptors^{10,11} (Fig. 4e). Together with the defects that we observed in IRF5 activation, this strongly suggested an analogous role for TASL as an innate immune adaptor for IRF5 that acts downstream of endolysosomal TLRs. Indeed, immunoprecipitation of the SLC15A4–TASL complex revealed CpG-induced recruitment of IRF5 (Fig. 4f–h, Extended Data Fig. 9g). Binding was lost both upon TASL knockout or use of the SLC15A4(E465K) mutant, which is deficient in TASL binding (Fig. 4f–h). Detailed mutagenesis of the TASL pLxIS motif demonstrated a strong functional analogy with the IRF3 adaptors STING, MAVS and TRIF. Mutation of the pLxIS motif or the core serine S294 abrogated TASL function, whereas a phosphomimetic S294D substitution retained detectable activity (Fig. 4i, j). As in STING and MAVS¹⁰, mutation of

Data Fig. 9h–k). TASL functionality was retained when its pLxIS motif was substituted with the corresponding sequences of IRF3 adaptors or of IRF3 itself, whereas a mutant containing the motif from IRF5 was less active (Extended Data Fig. 9l, m). The functional mutants activated IRF5 and not IRF3, which indicates that other structural determinants confer IRF specificity (Extended Data Fig. 9l).

To understand how TLR engagement signals to the SLC15A4–TASL complex and leads to TASL pLxIS-dependent IRF5 activation, we investigated kinases associated with this pathway. TBK1, IKK ϵ and IKK β have previously been shown to be involved in IRF3 adaptor phosphorylation and activation of IRF3 and/or IRF5^{10,35,45,46}. Similar to previous results on IRF5³⁵, co-expression of TASL with these kinases—but not kinase-dead mutants—resulted in a migratory shift indicative of hyperphosphorylation (Extended Data Fig. 10a). Treatment with the IKK β inhibitor TPCA-1 blocked R848-induced responses, IRF5 recruitment and activation, whereas the TBK1 and IKK ϵ inhibitor BX795 was less effective (Extended Data Fig. 10b–d). Furthermore, loss of IKK β as well as of the upstream kinases IRAK4 and TAK1 abrogated these responses, whereas TBK1, IKK ϵ or IKK α inactivation had no or only partial effects (Extended Data Fig. 10e–g). These data support a central role for IKK β in TASL-dependent activation of IRF5.

The work presented here identifies a complex between the previously uncharacterized protein TASL and SLC15A4—both genetically associated with SLE^{4–6}—as a functional module that is required for endolysosomal TLR signalling (Fig. 4k). In contrast to what might have been anticipated for an endolysosomally localized member of a transporter family, deficiency of neither SLC15A4 nor its partner TASL affected ligand–receptor engagement. Instead, loss of their function selectively impaired IRF5 signalling, whereas NF- κ B or MAPK induction was unaltered. The identification of a functional pLxIS motif in TASL revealed its role as innate immune adaptor that is required for recruitment and activation of IRF5 by TLR7, TLR8 and TLR9, in mechanistic analogy to IRF3 and its three adaptors STING, MAVS and TRIF¹⁰ (Fig. 4k, Extended Data Fig. 10h). In light of these findings, our study provides a molecular explanation for the involvement of TASL and SLC15A4 in SLE, defining the role of these interacting proteins in connecting endolysosomal TLRs to IRF5 (both established factors in autoimmune disease) and highlighting the potential of targeting the complex in pharmacological interventions.

Online content

Any methods, additional references, Nature Research reporting summaries, source data, extended data, supplementary information, acknowledgements, peer review information; details of author contributions and competing interests; and statements of data and code availability are available at <https://doi.org/10.1038/s41586-020-2282-0>.

- Blasius, A. L. & Beutler, B. Intracellular Toll-like receptors. *Immunity* **32**, 305–315 (2010).
- Pelka, K., Shibata, T., Miyake, K. & Latz, E. Nucleic acid-sensing TLRs and autoimmunity: novel insights from structural and cell biology. *Immunol. Rev.* **269**, 60–75 (2016).
- Kawai, T. & Akira, S. The role of pattern-recognition receptors in innate immunity: update on Toll-like receptors. *Nat. Immunol.* **11**, 373–384 (2010).
- Bentham, J. et al. Genetic association analyses implicate aberrant regulation of innate and adaptive immunity genes in the pathogenesis of systemic lupus erythematosus. *Nat. Genet.* **47**, 1457–1464 (2015).
- Odams, C. A. et al. Interferon inducible X-linked gene *CXorf21* may contribute to sexual dimorphism in systemic lupus erythematosus. *Nat. Commun.* **10**, 2164 (2019).
- Han, J. W. et al. Genome-wide association study in a Chinese Han population identifies nine new susceptibility loci for systemic lupus erythematosus. *Nat. Genet.* **41**, 1234–1237 (2009).
- Blasius, A. L. et al. SLC15a4, AP-3, and Hermansky–Pudlak syndrome proteins are required for Toll-like receptor signaling in plasmacytoid dendritic cells. *Proc. Natl Acad. Sci. USA* **107**, 19973–19978 (2010).
- Sasawatari, S. et al. The solute carrier family 15A4 regulates TLR9 and NOD1 functions in the innate immune system and promotes colitis in mice. *Gastroenterology* **140**, 1513–1525 (2011).
- Kobayashi, T. et al. The histidine transporter SLC15A4 coordinates mTOR-dependent inflammatory responses and pathogenic antibody production. *Immunity* **41**, 375–388 (2014).
- Liu, S. et al. Phosphorylation of innate immune adaptor proteins MAVS, STING, and TRIF induces IRF3 activation. *Science* **347**, aaa2630 (2015).
- Zhao, B. et al. Structural basis for concerted recruitment and activation of IRF-3 by innate immune adaptor proteins. *Proc. Natl Acad. Sci. USA* **113**, E3403–E3412 (2016).
- Tsokos, G. C. Systemic lupus erythematosus. *N. Engl. J. Med.* **365**, 2110–2121 (2011).
- Moulton, V. R. et al. Pathogenesis of human systemic lupus erythematosus: a cellular perspective. *Trends Mol. Med.* **23**, 615–635 (2017).
- Tsokos, G. C., Lo, M. S., Costa Reis, P. & Sullivan, K. E. New insights into the immunopathogenesis of systemic lupus erythematosus. *Nat. Rev. Rheumatol.* **12**, 716–730 (2016).
- Kieser, K. J. & Kagan, J. C. Multi-receptor detection of individual bacterial products by the innate immune system. *Nat. Rev. Immunol.* **17**, 376–390 (2017).
- Janeway, C. A. Jr & Medzhitov, R. Innate immune recognition. *Annu. Rev. Immunol.* **20**, 197–216 (2002).
- O'Neill, L. A., Goldenbogen, D. & Bowie, A. G. The history of Toll-like receptors – redefining innate immunity. *Nat. Rev. Immunol.* **13**, 453–460 (2013).
- Blasius, A. L., Krebs, P., Sullivan, B. M., Oldstone, M. B. & Popkin, D. L. SLC15a4, a gene required for pDC sensing of TLR ligands, is required to control persistent viral infection. *PLoS Pathog.* **8**, e1002915 (2012).
- Baccala, R. et al. Essential requirement for IRF8 and SLC15A4 implicates plasmacytoid dendritic cells in the pathogenesis of lupus. *Proc. Natl Acad. Sci. USA* **110**, 2940–2945 (2013).
- Dosenovic, P. et al. SLC15a4 function is required for intact class switch recombination to IgG2c in response to TLR9 stimulation. *Immunol. Cell Biol.* **93**, 136–146 (2015).
- Griffith, A. D. et al. A requirement for SLC15a4 in imiquimod-induced systemic inflammation and psoriasiform inflammation in mice. *Sci. Rep.* **8**, 14451 (2018).
- Langeveld, C. D. et al. Transancestral mapping and genetic load in systemic lupus erythematosus. *Nat. Commun.* **8**, 16021 (2017).
- Pollard, K. M. et al. Induction of systemic autoimmunity by a xenobiotic requires endosomal TLR trafficking and signaling from the late endosome and endolysosome but not type I IFN. *J. Immunol.* **199**, 3739–3747 (2017).
- Rebsamen, M. et al. SLC38A9 is a component of the lysosomal amino acid sensing machinery that controls mTORC1. *Nature* **519**, 477–481 (2015).
- Pichlmair, A. et al. Viral immune modulators perturb the human molecular network by common and unique strategies. *Nature* **487**, 486–490 (2012).
- Nakamura, N. et al. Endosomes are specialized platforms for bacterial sensing and NOD2 signalling. *Nature* **509**, 240–244 (2014).
- Harris, V. M., Harley, I. T. W., Kurien, B. T., Koelsch, K. A. & Scofield, R. H. Lysosomal pH is regulated in a sex dependent manner in immune cells expressing *CXorf21*. *Front. Immunol.* **10**, 578 (2019).
- Libert, C., Dejager, L. & Pinheiro, I. The X chromosome in immune functions: when a chromosome makes the difference. *Nat. Rev. Immunol.* **10**, 594–604 (2010).
- Klijn, C. et al. A comprehensive transcriptional portrait of human cancer cell lines. *Nat. Biotechnol.* **33**, 306–312 (2015).
- FANTOM Consortium and the RIKEN PMI and CLST. A promoter-level mammalian expression atlas. *Nature* **507**, 462–470 (2014).
- Arazi, A. et al. The immune cell landscape in kidneys of patients with lupus nephritis. *Nat. Immunol.* **20**, 902–914 (2019).
- Muskardin, T. L. W. & Niewold, T. B. Type I interferon in rheumatic diseases. *Nat. Rev. Rheumatol.* **14**, 214–228 (2018).
- Hu, Y., Song, F., Jiang, H., Nuñez, G. & Smith, D. E. SLC15A2 and SLC15A4 mediate the transport of bacterially derived di/tripeptides to enhance the nucleotide-binding oligomerization domain-dependent immune response in mouse bone marrow-derived macrophages. *J. Immunol.* **201**, 652–662 (2018).
- Lee, J. et al. pH-dependent internalization of muramyl peptides from early endosomes enables Nod1 and Nod2 signaling. *J. Biol. Chem.* **284**, 23818–23829 (2009).
- Chow, K. T. et al. Differential and overlapping immune programs regulated by IRF3 and IRF5 in plasmacytoid dendritic cells. *J. Immunol.* **201**, 3036–3050 (2018).
- Graham, R. R. et al. A common haplotype of interferon regulatory factor 5 (*IRF5*) regulates splicing and expression and is associated with increased risk of systemic lupus erythematosus. *Nat. Genet.* **38**, 550–555 (2006).
- Schoenemeyer, A. et al. The interferon regulatory factor, IRF5, is a central mediator of Toll-like receptor 7 signaling. *J. Biol. Chem.* **280**, 17005–17012 (2005).
- Takaoka, A. et al. Integral role of IRF-5 in the gene induction programme activated by Toll-like receptors. *Nature* **434**, 243–249 (2005).
- Eames, H. L., Corbin, A. L. & Udalova, I. A. Interferon regulatory factor 5 in human autoimmunity and murine models of autoimmune disease. *Transl. Res.* **167**, 167–182 (2016).
- Maeda, T. et al. A novel plasmacytoid dendritic cell line, CAL-1, established from a patient with blastic natural killer cell lymphoma. *Int. J. Hematol.* **81**, 148–154 (2005).
- Steinhagen, F. et al. IRF-5 and NF- κ B p50 co-regulate IFN- β and IL-6 expression in TLR9-stimulated human plasmacytoid dendritic cells. *Eur. J. Immunol.* **43**, 1896–1906 (2013).
- Newstead, S. Recent advances in understanding proton coupled peptide transport via the POT family. *Curr. Opin. Struct. Biol.* **45**, 17–24 (2017).
- Kelley, L. A., Mezulis, S., Yates, C. M., Wass, M. N. & Sternberg, M. J. The Phyre2 web portal for protein modeling, prediction and analysis. *Nat. Protocols* **10**, 845–858 (2015).
- Chen, W. et al. Insights into interferon regulatory factor activation from the crystal structure of dimeric IRF5. *Nat. Struct. Mol. Biol.* **15**, 1213–1220 (2008).
- Lopez-Pelaez, M. et al. Protein kinase IKK β -catalyzed phosphorylation of IRF5 at Ser462 induces its dimerization and nuclear translocation in myeloid cells. *Proc. Natl Acad. Sci. USA* **111**, 17432–17437 (2014).
- Ren, J., Chen, X. & Chen, Z. J. IKK β is an IRF5 kinase that instigates inflammation. *Proc. Natl Acad. Sci. USA* **111**, 17438–17443 (2014).
- Doench, J. G. et al. Optimized sgRNA design to maximize activity and minimize off-target effects of CRISPR–Cas9. *Nat. Biotechnol.* **34**, 184–191 (2016).

48. Bigenzahn, J. W. et al. LZTR1 is a regulator of RAS ubiquitination and signaling. *Science* **362**, 1171–1177 (2018).
49. Brinkman, E. K., Chen, T., Amendola, M. & van Steensel, B. Easy quantitative assessment of genome editing by sequence trace decomposition. *Nucleic Acids Res.* **42**, e168 (2014).
50. Troegeler, A. et al. An efficient siRNA-mediated gene silencing in primary human monocytes, dendritic cells and macrophages. *Immunol. Cell Biol.* **92**, 699–708 (2014).
51. Rudashevskaya, E. L. et al. A method to resolve the composition of heterogeneous affinity-purified protein complexes assembled around a common protein by chemical cross-linking, gel electrophoresis and mass spectrometry. *Nat. Protocols* **8**, 75–97 (2013).
52. Varjosalo, M. et al. Interlaboratory reproducibility of large-scale human protein-complex analysis by standardized AP-MS. *Nat. Methods* **10**, 307–314 (2013).
53. Olsen, J. V. et al. Parts per million mass accuracy on an Orbitrap mass spectrometer via lock mass injection into a C-trap. *Mol. Cell. Proteomics* **4**, 2010–2021 (2005).
54. Chambers, M. C. et al. A cross-platform toolkit for mass spectrometry and proteomics. *Nat. Biotechnol.* **30**, 918–920 (2012).
55. Apweiler, R. et al. UniProt: the Universal Protein knowledgebase. *Nucleic Acids Res.* **32**, D115–D119 (2004).
56. Cox, J. & Mann, M. MaxQuant enables high peptide identification rates, individualized p.p.b.-range mass accuracies and proteome-wide protein quantification. *Nat. Biotechnol.* **26**, 1367–1372 (2008).
57. Wright, J. C. & Choudhary, J. S. DecoyPyrat: fast non-redundant hybrid decoy sequence generation for large scale proteomics. *J. Proteomics Bioinform.* **9**, 176–180 (2016).
58. Kim, S. & Pevzner, P. A. M. S.-G. F. MS-GF+ makes progress towards a universal database search tool for proteomics. *Nat. Commun.* **5**, 5277 (2014).
59. Elias, J. E. & Gygi, S. P. Target-decoy search strategy for increased confidence in large-scale protein identifications by mass spectrometry. *Nat. Methods* **4**, 207–214 (2007).
60. The, M., MacCoss, M. J., Noble, W. S. & Käll, L. Fast and accurate protein false discovery rates on large-scale proteomics data sets with percolator 3.0. *J. Am. Soc. Mass Spectrom.* **27**, 1719–1727 (2016).
61. Choi, H. et al. Analyzing protein-protein interactions from affinity purification-mass spectrometry data with SAINT. *Curr. Protoc. Bioinformatics* **39**, 8.15.1–8.15.23 (2012).
62. Shannon, P. et al. Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Res.* **13**, 2498–2504 (2003).
63. Mellacheruvu, D. et al. The CRAPome: a contaminant repository for affinity purification-mass spectrometry data. *Nat. Methods* **10**, 730–736 (2013).
64. UniProt Consortium. UniProt: a worldwide hub of protein knowledge. *Nucleic Acids Res.* **47** (D1), D506–D515 (2019).
65. Larkin, M. A. et al. Clustal W and Clustal X version 2.0. *Bioinformatics* **23**, 2947–2948 (2007).
66. Drozdetskiy, A., Cole, C., Procter, J. & Barton, G. J. JPred4: a protein secondary structure prediction server. *Nucleic Acids Res.* **43**, W389–W394 (2015).
67. Dobin, A. et al. STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* **29**, 15–21 (2013).
68. Derr, A. et al. End Sequence Analysis Toolkit (ESAT) expands the extractable information from single-cell RNA-seq data. *Genome Res.* **26**, 1397–1410 (2016).
69. Love, M. I., Huber, W. & Anders, S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* **15**, 550 (2014).
70. Marbach, D. et al. Tissue-specific regulatory circuits reveal variable modular perturbations across complex diseases. *Nat. Methods* **13**, 366–370 (2016).
71. Hochberg, Y. & Benjamini, Y. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J. R. Stat. Soc.* **57**, 289–300 (1995).
72. Huang, W., Sherman, B. T. & Lempicki, R. A. Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nat. Protocols* **4**, 44–57 (2009).
73. Blaszczyk, M., Ciemny, M. P., Kolinski, A., Kurcinski, M. & Kmiecik, S. Protein-peptide docking using CABS-dock and contact information. *Brief. Bioinform.* **20**, 2299–2305 (2019).
74. London, N., Raveh, B., Cohen, E., Fathi, G. & Schueler-Furman, O. Rosetta FlexPepDock web server—high resolution modeling of peptide–protein interactions. *Nucleic Acids Res.* **39**, W249–W253 (2011).
75. Colas, C. et al. An improved flow cytometry assay to monitor phagosome acidification. *J. Immunol. Methods* **412**, 1–13 (2014).
76. Carpenter, A. E. et al. CellProfiler: image analysis software for identifying and quantifying cell phenotypes. *Genome Biol.* **7**, R100 (2006).

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

© The Author(s), under exclusive licence to Springer Nature Limited 2020

Methods

No statistical methods were used to predetermine sample size. The experiments were not randomized and investigators were not blinded to allocation during experiments and outcome assessment.

Antibodies and reagents

HA (no. 3724), V5 (no. 13202), GFP (no. 2956), RAGA (no. 4357), TLR9 (no. 5845), TAK1 (no. 4505), IRAK4 (no. 4363), IKK α (no. 2682), IKK β (no. 8943), TBK1 (no. 51872), IKK ϵ (no. 2905), IRF3 (no. 11904), IRF7 (no. 4920), phospho-IKK α / β Ser176/180 (no. 2697), phospho-IkB α Ser32/36 (no. 9246), phospho-NF- κ B p65 Ser536 (no. 3033), phospho-p38 MAPK Thr180/Tyr182 (no. 9211), phospho-SAPK/JNK Thr183/Tyr185 (no. 4668) and phospho-STAT1 Y701 (no. 7649) antibodies were from Cell Signaling; anti-CXorf21 (TASL) (HPA001185), SLC38A9 (HPA043785, for IP) and anti-MYC (C3956) from Sigma; anti-SLC15A4 (BMP055, for IP) and TLR8 (pd047) from MBL; anti-GAPDH (sc-365062), anti-IkB α (sc-371), anti-LAMP2 (sc-18822), anti-HA (sc-805, for IP) and anti-STAT1 α (sc-417) from Santa Cruz; anti-actin (AAN01-A) from Cytoskeleton, anti-tubulin (ab7291), anti-IRF5 (ab181553) and anti-LAMP1 (ab25630) from Abcam. Custom rabbit anti-SLC15A4 antibodies raised against the N terminus were generated with Genscript. Specificity was validated by western blot in SLC15A4-knockout and -overexpressing cells (Fig. 2a, Extended Data Fig. 3g). Fluorescently labelled anti-PD-L1-APC (17-5983-42) was from Thermo Fisher Scientific, Alexa-Fluor-488-coupled anti-mouse (A11001) and Alexa-Fluor-594-coupled anti-rabbit (A11012) were from Life Technologies. Anti-HA and anti-V5 agarose beads were from Sigma, Protein G sepharose beads were from GE Healthcare. LysoTracker Red DND-99 and LysoSensor Green DND-189 were from Thermo Fisher Scientific. R848, CL075, ssR40, ssR41, C12-iE-DAP, MDP, murabutide, ultrapure LPS (*Escherichia coli* 0111:B4), Pam3CSK4, cGAMP, flagellin from *Salmonella typhimurium*, unlabelled or FITC-labelled CpG-A (ODN2216), CpG-B (ODN2006) and BX795 were from Invivogen. TPCA-1 and PMA from Sigma. Recombinant human M-CSF, GM-CSF, IL-4, interferon- β , and interferon- γ were from Peprotech. λ phosphatase and PNGase F were from NEB.

Cell culture

HEK293T cells and THP1 cells were purchased from ATCC, THP1 DUAL reporter cell lines from Invivogen. CAL-1 cells⁴⁰ were provided by T. Maeda and KBM-7 cells by T. Brummelkamp. Cells—except for CAL-1—were authenticated by short tandem repeat profiling and regularly tested for mycoplasma contamination. HEK293T cells were cultured in DMEM, THP1 and CAL-1 in RPMI and KBM-7 cells in IMDM, supplemented with 10% (v/v) FBS and antibiotics (100 U/ml penicillin, 100 mg/ml streptomycin), all from Gibco. Cells were incubated at 37 °C in 5% CO₂.

For differentiation of primary monocyte-derived macrophages and dendritic cells, CD14⁺ monocytes were seeded in 6-well plates at a concentration of 1.5×10^6 cells in 2 ml per well RPMI medium supplemented with 10% (v/v) FBS and antibiotics (100 U/ml penicillin, 100 mg/ml streptomycin). Monocyte-derived macrophages were generated by stimulation for 1 week with 100 ng/ml M-CSF, dendritic cells by stimulation with 200 ng/ml GM-CSF and 50 ng/ml IL-4.

Plasmids and siRNAs

CRISPR–Cas9-based knockout cell line generation was performed using pLentiCRISPRv2 (Addgene plasmid no. 52961). sgRNAs were designed using the Broad Institute sgRNA design tool (<https://portals.broadinstitute.org/gpp/public/analysis-tools/sgRNA-design>)⁴⁷; the control sgRNA targeting *Renilla luciferase* (sgRen) has previously been described⁴⁸. Editing efficiencies for sgRNAs targeting *SLC15A2* and *SLC15A3* were determined by ‘Tracking of Indels by Decomposition’ (TIDE)⁴⁹. Cloned oligonucleotides were as follows (5′ to 3′ orientation): *SLC15A2* sgRNA no. 1, forward (F): CACCGGATATAAGGAATAGTACCC, reverse (R): AAACGGGTACTATTCCTTTATATCC; *SLC15A2* sgRNA no. 2,

F: CACCGACTGAGCATTGCCTTCATTG, R: AAACCAATGAAGGCAATG CTCAGTC; *SLC15A2* sgRNA no. 3, F: CACCGAGGAGGCATCAAACCCCTG TG, R: AAACCACAGGGTTTGATGCCTCCTC; *SLC15A3* sgRNA no. 1, F: CACCGTTGGCGATGTCCTCTTGCG, R: AAACCGCAAGAGGACATCGCCA ACC; *SLC15A3* sgRNA no. 2, F: CACCGGAGAGCGAGCTTAAGCATAG, R: AAACCTATGCTTAAGCTCGCTCTCC; *SLC15A3* sgRNA no. 3, F: CACCGG CAGCAGACAGCACAGCACCC, R: AAACGGGTGCTGTGCTGTGCTGCTGCC; *SLC15A4* sgRNA no. 1, F: CACCGGGAGCGATCCTGTCTGTAGG, R: AAACCC TAACGACAGGATCGCTCCC; *SLC15A4* sgRNA no. 2, F: CACCGTATTAC AACCCTCCTCACA, R: AAACCTGTGAGGAGTGGTTGTAATAC; *CXorf21* sgRNA no. 1, F: CACCGGTAGAAATGGAATCCTCCAT, R: AAACATGGAGG ATTCCATTCTACC; *CXorf21* sgRNA no. 2, F: CACCGCTGAATTAATGGC CATCACC, R: AAACGGTGATGGCCATTAATTCAGC; *IRF3* sgRNA no. 1, F: CACCGGAGGTGACAGCCTTCTACCG, R: AAACCGGTAGAAGGCTGTCA CCTCC; *IRF3* sgRNA no. 2, F: CACCGCCACTGGTGCATATGTTCCC, R: AAACGGGAACATATGCACCATGGGC; *IRF5* sgRNA no. 1, F: CACCGAGGGCTTCAATGGGTCAACG, R: AAACCGTTGACCCATTGAAGCC CTC; *IRF5* sgRNA no. 2, F: CACCGATGAAGCCGATCCGGCCAAAG, R: AAACCTTGGCCGGATCGGCTTCATC; *IRF7* sgRNA no. 1, F: CACCGG ATGCACTCACCTTGACCCG, R: AAACCGGTGCAAGGTGAGTGCATCC; *IRF7* sgRNA no. 2, F: CACCGGGCAGATCCAGTCCCAACCA, R: AAACCTGG TTGGGACTGGATCTGCCC; *TLR8* sgRNA, F: CACCGACAGGAAGTCCCC CAAACGG, R: AAACCGTTTGGGGAACTTCTGTGTC; *CHUK* sgRNA no. 1, F: CACCGAAAGCTCCAATAATCAACAG, R: AAACCTGTGATTATTGG AGCTTTC; *CHUK* sgRNA no. 2, F: CACCGTATACAGCTGCGTAAAGTGT, R: AAACACACTTTACGCAGCTGTATAC; *CHUK* sgRNA no. 3, F: CACCGT AGTTTAGTAGTAGAACCAC, R: AAACCTGGTTCTACTACTAACTAC; *IKBKB* sgRNA no. 1, F: CACCGGCCATGGAGTACTGCCAAGG, R: AAACC CTTGGCAGTACTCCATGGCC; *IKBKB* sgRNA no. 2, F: CACCGCAGCC ATTGGGCCCATACGT, R: AAACACGTATGGGCCCAATGGCTGC; *IKBKB* sgRNA no. 3, F: CACCGTATTGACCTAGGATATGCCA, R: AAACCTGGCATA TCCTAGGTCAATAC; *IKBKB* sgRNA no. 4, F: CACCGGAAGCCCGTGATG CACTCAA, R: AAACCTGAGTGCATCACGGGCTTCC; *IKBKE* sgRNA no. 1, F: CACCGTCAACACTACCAGCTACCTG, R: AAACAGGTAGCTGGT AGTGTGAC; *IKBKE* sgRNA no. 2, F: CACCGCGTGACAAGCAGACG ACCAGT, R: AAACCACTGGTCTGTCTGTGACCG; *IKBKE* sgRNA no. 3, F: CACCGATGATCTCCTTGTTCGCCG, R: AAACCGGCGGAACAAGGAG ATCATC; *IRAK4* sgRNA no. 1, F: CACCGCTACGTAAATAACACAACCTG, R: AAACCACTGTGTTATTTACGTAGC; *IRAK4* sgRNA no. 2, F: CACC GGGCACCACAAATTGCACAGT, R: AAACACTGTGCAATTTGTGGTG CCC; *IRAK4* sgRNA no. 3, F: CACCGCATCTCATGTGCCAAGAAAG, R: AAACCTTTCTTGGCACATGAGATGC; *IRAK4* sgRNA no. 4, F: CACCGTG TAAACATATACTAAGCAG, R: AAACCTGCTTAGTATATGTTTACAC; *MAP3K7* sgRNA no. 1, F: CACCGACCCAAAGCGCTAATTCACA, R: AAACCTG TGAATTAGCGCTTTGGGTC; *MAP3K7* sgRNA no. 2, F: CACCGAATATTA GGATGGTTACAC, R: AAACGTGTGAACCATCCTAATATTTC; *MAP3K7* sgRNA no. 3, F: CACCGCACACATGACCAATAACAAG, R: AAACCTTGT ATTGGTCATGTGTGC; *TBK1* sgRNA no. 1, F: CACCGTCCAGTTATGATTTA GACG, R: AAACCGTCTAAATCATAACGTGGAC; *TBK1* sgRNA no. 2, F: CACCGAATCAAGAACTTATCTACGA, R: AAACCTCGTAGATAAGTTCTTG ATTC; and *TBK1* sgRNA no. 3, F: CACCGAAATATCATGCGTGTATAG, R: AAACCTATAACACGCATGATATTC.

Codon-optimized cDNAs for human SLC15A3, human and mouse SLC15A4, SLC15A3–SLC15A4 swap mutants, and human wild type and scanning mutants of TASL were obtained from Genscript. A template for cloning of mouse SLC15A3 was obtained from the Harvard Plasmid Repository (clone identifier: MmCD00319552), a template for cloning human TLR9 (pcDNA3-TLR9-YFP) was from Addgene (plasmid no. 13642). cDNAs were subcloned to pDONR201 (Invitrogen) via Gateway cloning. Gateway donor plasmids for GFP and human SLC38A9 have previously been described²⁴. Deletion and point mutants were generated by PCR or Q5 mutagenesis (NEB). All cDNAs were verified by sequencing and shuttled to Gateway destination vectors for untagged or N- or C-terminal Strep-HA-tagged (SH), V5 or MYC-tagged expression. Rescue experiments were performed using codon-optimized cDNAs resistant

Article

to sgRNAs targeting the endogenous genes. For lentiviral transduction, cDNAs were shuttled to pRRRL-based lentiviral expression plasmids and a previously described selectable resistance cassette⁴⁸. Lentiviral packaging plasmids psPAX2 (plasmid no. 12260) and pMD2.G (plasmid no. 12259) were obtained from Addgene. Non-targeting control and human *SLC15A4*, *CXorf21* and *MYD88*-specific ON-TARGETplus SMARTpool siRNAs were obtained from Dharmacon. Non-targeting pool, control siRNA (cat.: D-001810-10-05): siRNA no. 1: 5'-UGGUUUACAUGUCGACUAA-3', siRNA no. 2: 5'-UGGUUUACAUGUUGUGUGA-3', siRNA no. 3: 5'-UGGUUUACAUGUUUUCUGA-3', siRNA no. 4: 5'-UGGUUUACAUGUUUCCUA-3'). Human *SLC15A4* (cat.: L-007401-02-0005): siRNA no. 1: 5'-CGACCAGGUUAAAGAUCGA-3', siRNA no. 2: 5'-GAAGC GAAGUGGAGAGCGC-3', siRNA no. 3: 5'-CCUGAGGCCAUGUGCGGUU-3', siRNA no. 4: 5'-GCAUUAACUGGGAGCGAU-3'). Human *CXorf21* (*TASL*) (cat.: L-014600-02-0005): siRNA no. 1: 5'-GCAGAAGGUUGU GGAGUUA-3', siRNA no. 2: 5'-CAAUGUAAAUCCAAGAGA-3', siRNA no. 3: 5'-GGACUUGAGUACUGGAAUG-3', siRNA no. 4: 5'-CC AUUAAUUCAGUGACAAC-3'). Human *MYD88* (cat.: L-004769-00-0005): siRNA no. 1: 5'-CGACUGAAGUUGUGUGUGU-3', siRNA no. 2: 5'-GCUAGUGAGCUCAUCGAAA-3', siRNA no. 3: 5'-GCAUAUGC CUGAGCGUUUC-3', siRNA no. 4: 5'-GCACCGUGUCUGGUCUAU-3'.

Lentiviral gene transduction

For lentiviral gene transduction, HEK293T cells were transfected with the respective lentiviral vectors and packaging plasmids psPAX2 and pMD2.G using Polyfect (Qiagen) or PEI (Sigma). Twenty-four hours later, medium was exchanged to RPMI, supplemented with 10% (v/v) FBS and antibiotics (100 U/ml penicillin, 100 mg/ml streptomycin). Forty-eight hours after transfection, cell supernatants were collected, filtered through 0.45-µm polyethersulfone filters (GE Healthcare) and supplemented with 8 µg/ml protamine sulfate (Sigma). Cells were infected by spinfection (2,000 rpm, 45 min, room temperature). Twenty-four hours after infection, medium was changed; 48 h after infection, cells were selected with the respective antibiotics.

Cell lysis, western blotting and co-immunoprecipitation

Cells were lysed in RIPA (25 mM Tris, 150 mM NaCl, 0.5% NP-40, 0.5% deoxycholate (w/v) and 0.1% SDS (w/v), pH 7.4) or EIA (50 mM HEPES, 250 mM NaCl, 5 mM EDTA, 1% NP-40, pH 7.4) lysis buffer supplemented with Roche EDTA-free protease inhibitor cocktail (1 tablet per 50 ml) for 10 min on ice. Where phosphoproteins were analysed, lysis buffer was supplemented with Halt phosphatase inhibitor cocktail (Thermo Fisher Scientific). Lysates were cleared by centrifugation at 13,000 rpm, 10 min, 4 °C and normalized by Bradford protein assay (Bio-Rad) or BCA (Thermo Fisher Scientific) using BSA as standard. Typically, 20 µg protein per sample was resolved by regular or 20 µM Phos-tag-containing (WAKO Chemicals) SDS-PAGE and blotted to nitrocellulose membranes. Membranes were blocked with 5% non-fat dry milk in TBST and probed with the indicated antibodies. Binding was detected with horseradish-peroxidase-conjugated secondary antibodies using the ECL western blotting system (Thermo Fisher Scientific). In experiments in which multiple antibodies were used, equal amounts of samples were loaded on multiple SDS-PAGE gels and western blots sequentially probed with a maximum of two antibodies. For immunoprecipitation experiments of overexpressed proteins, cells were lysed in EIA buffer. Whole-cell lysate was removed as input, and the rest was subjected to immunoprecipitation using equilibrated anti-HA or V5 agarose beads (Sigma) overnight at 4 °C. Beads were washed three times with EIA buffer and eluted with SDS sample buffer. In experiments monitoring co-immunoprecipitation of IRF5, the second wash step was performed with EIA buffer containing a higher NaCl concentration (500 mM). Samples were analysed by western blot as described. For immunoprecipitation of endogenous proteins, 1.5×10^7 cells per condition were lysed in 500 µl EIA buffer. Forty microlitres whole-cell lysate was removed as input, and the rest was subjected first to a pre-clearing step on Sepharose 6

beads (Sigma) (40 min with rotation, 4 °C) and then to immunoprecipitation using 20 µl equilibrated protein G sepharose (GE Healthcare) and primary antibody (SLC15A4 BMP055 MBL; SLC38A9 HPA043785 Sigma; HA sc-805 Santa Cruz) (overnight with rotation, 4 °C). Beads were washed three times with EIA buffer and eluted with 60 µl SDS sample buffer.

PNGase F treatment

Cells were lysed in EIA buffer. Per sample, 20 µl cleared lysate was either incubated without or with 1–2 µl (500–1,000 U) PNGase F (NEB) for 30 min at 37 °C. Samples were analysed by western blotting.

λ phosphatase treatment

Immunoprecipitates were generated as described in 'Cell lysis, western blotting and co-immunoprecipitation'. Washed beads were resuspended in 60 µl NEB PMP buffer + 1 mM MnCl₂, split in two and incubated or not with 1 µl (400 U) λ phosphatase (NEB) for 30 min at 30 °C. Samples were analysed by western blot.

Enzyme-linked immunosorbent assay

All enzyme-linked immunosorbent assay (ELISA) experiments were carried out using diluted cell culture supernatants according to the manufacturer's instructions. ELISA kits for human TNF (no. 88-7346-88) and IL-8 (no. 88-8086-88) were from Invitrogen, human IL-6 (no. 88-7066-88) from eBioscience; ELISA kits for human CCL2 (no. DY279) and CCL5 (no. DY278) were from R&D Systems; and for human interferon-β from PBL Assay Science (41410-1).

THP1 DUAL cell reporter assay

THP1 DUAL cells (1×10^5 cells per 96 well) were stimulated as indicated for 20–24 h. Cell culture supernatants were collected, cleared of residual cells by centrifugation and analysed for NF-κB and ISRE reporter activity according to the manufacturer's instructions.

Flow cytometry

For PD-L1, cells were stained with APC-conjugated anti-PD-L1 antibodies (17-5983-42, Thermo Fisher Scientific) according to the manufacturer's instructions. For uptake assays of FITC-labelled CpG, cells were incubated with 1 µM CpG-A (ODN2216) or CpG-B (ODN2006) for 0–120 min. Cells were washed with PBS and analysed immediately by flow cytometry. Data were acquired on a BD FACSCalibur (BD Biosciences) and analysed using FlowJo software (version 10).

Confocal microscopy

For staining of fixed cells, 1×10^5 cells were seeded in 24-well plates on cover slips and treated with 10 nM PMA overnight to induce adherence. Cells were fixed for 10 min with 4% formaldehyde in PBS and permeabilized and blocked with 0.3% saponin (Sigma) and 10% FBS in PBS for 1 h. Cells were stained overnight at 4 °C with the indicated primary antibodies (rabbit anti-HA: 1:400, mouse anti-LAMP1: 1:200) in blocking solution. Cells were washed 3 times in blocking solution and stained for 1 h with fluorescently labelled anti-rabbit and anti-mouse secondary antibodies (1:400) at room temperature. Cells were washed three times in blocking solution and once in PBS. Nuclear counterstaining was performed with DAPI (Thermo Fisher Scientific), diluted 1:1,000 in PBS and cover glasses were mounted onto microscope slides using ProLong Gold (Thermo Fisher Scientific) antifade reagent. For live-cell imaging of TASL-GFP-expressing THP1 cells, cells were stained with LysoTracker Red DND-99 (1:10,000) and Hoechst 33342 (1:1,000, Thermo Fisher Scientific) for 30 min and washed with PBS. Images were acquired on a confocal laser scanning microscope (Zeiss LSM 780, Carl Zeiss) and analysed using ZEN 2.3 (Carl Zeiss).

siRNA knockdown in human primary monocytes

Peripheral blood mononuclear cells from healthy donors were obtained by density gradient centrifugation of buffy coat material obtained from

the Austrian Red Cross with Lymphoprep (Stem Cell Technologies). CD14⁺ monocytes were purified from peripheral blood mononuclear cells using CD14-specific MACS immunomagnetic beads (Miltenyi) according to the manufacturer's instructions. siRNAs were transfected according to a modified version of a previously described protocol⁵⁰. Lipid-siRNA complexes were prepared by combining 15 μ l siRNA (20 μ M stock) with 470 μ l non-supplemented RPMI medium and addition of 15 μ l HiPerfect transfection reagent (Qiagen). After 15–20 min, complexes were transferred to 6-well plates and combined with 1.5×10^6 monocytes per well in 1 ml RPMI medium supplemented with 10% (v/v) FBS and antibiotics (100 U/ml penicillin, 100 mg/ml streptomycin). The next day, 1 ml supplemented RPMI medium was added to each well. Stimulation experiments and RNA isolation for analysis of knockdown efficiency were carried out 48 h after transfection. Cell supernatants were analysed for TNF secretion by ELISA. On the basis of the seven donors, the following differences were tested using a paired *t*-test (two-sided) on the log₂-transformed TNF concentrations, assuming normality: *SLC15A4* siRNA versus control siRNA (effect size: -0.69; 95% confidence interval: [-1.31, -0.07]; *t*-statistic: -2.72; *P*-value: 0.035; degrees of freedom: 6), *Cxorf21* siRNA versus control siRNA (effect size: -1.97; 95% confidence interval: [-2.63, -1.30]; *t*-statistic: -7.19; *P*-value: 3.66×10^{-4} ; degrees of freedom: 6), and *MYD88* siRNA versus control siRNA (effect size: -0.90; 95% confidence interval: [-1.73, -0.07]; *t*-statistic: -2.65; *P*-value: 0.038; degrees of freedom: 6).

Real-time PCR

RNA from monocytes was isolated 48 h after siRNA transfection using the Qiagen RNeasy Mini kit including DNase I digestion step. The reverse transcription was performed using RevertAid First Strand cDNA Synthesis Kit (Thermo Scientific) using oligo dT primers. Real-time PCR was performed using the SensiFAST SYBR Hi-ROX kit (Bioline) according to the manufacturer's instructions. The primers used are: *GAPDH*: 5'-CCTGACCTGCCGTCTAGAAA-3', R: 5'-CTCCGACGCCTGCTTCAC-3'; *SLC15A4* F: 5'-CGGATGGATGAGCAGTCACA-3', R: 5'-AGGAAAGCAGGAGGGTAGC-3'; *Cxorf21* F: 5'-GGAAAGAGCATTGGCTGGCTT-3', R: 5'-TTCTCACACTGACCTTCACTAACCA-3'; *MYD88* F: 5'-GAGCTCATCGAAAAGAGGTGC-3', R: 5'-GGAGAGAGGCTGAGTGCAA-3'. Samples were analysed on a LightCycler 480 (Roche) or Rotor Gene Q (Qiagen). Amplification on LightCycler 480 consisted of an initial incubation at 95 °C for 10 min, followed by 40 cycles of 95 °C for 5 s, 60 °C for 60 s and 72 °C for 6 s and a final cooling to 40 °C. Data were analysed and *C_t* values were calculated using LightCycler Software version 1.5 (Roche). Amplification on Rotor Gene Q consisted of an initial incubation at 95 °C for 10 min, followed by 40 cycles of 94 °C for 30 s, 60 °C for 15 s and 72 °C for 30 s and a final cooling to 25 °C. Data were analysed and *C_t* values were calculated using the Rotor Gene Series Software version 2.2.2 (Qiagen). Results were obtained using the 2^{- $\Delta\Delta C_t$} method, using *GAPDH* as reference.

Affinity purification and mass spectrometry

Affinity purifications and sample preparation for LC-MS/MS were performed as previously described^{24,51,52}. Two affinity purifications were performed as biological replicates, and cell lines expressing Strep-HA-tagged GFP were used as negative controls. LC-MS/MS was performed on the following instruments: hybrid linear trap quadrupole (LTQ) Orbitrap Velos, Q Exactive or Orbitrap Fusion Lumos Tribrid mass spectrometer (Thermo Fisher Scientific) coupled to either an Agilent 1200 (Agilent Biotechnologies) or Dionex U3000RSLC U/HPLC nano-flow system (Thermo Fisher Scientific) via Nanospray Flex Ion source interface. Tryptic peptides were loaded onto a trap column using 0.1% TFA as loading buffer. After loading, the trap column was switched in-line with a 75- μ m-inner-diameter analytical column (packed in-house with ReproSil-Pur 120 C18-AQ, 3 μ m, Dr Maisch). Mobile-phase A consisted of 0.4% formic acid in water and mobile-phase B of 0.4% formic acid in a mix of 90% acetonitrile and 9.6% water. The flow rate was set

to 230 nl/min and either a 30 min (3 to 36% solvent B within 30 min) or a 90 min (4 to 30% solvent B within 81 min, 30 to 65% solvent B within 8 min and, 65 to 100% solvent B within 1 min) gradient was applied, followed by flushing at 100% solvent B for 6 min before re-equilibration of the column material at 3% solvent B for 18 min before next injection. Mass spectrometry instruments were operated in data-dependent acquisition mode with top 10 or 15 most intense precursor ions selected for collision-induced dissociation in the linear ion trap (LTQ) or higher energy collision induced dissociation (HCD) in the HCD cell. MS1 spectra were acquired in the Orbitrap mass analyser at high resolution, and MS² fragment ion spectra were acquired either in the linear ion trap at low resolution (Velos) or in the Orbitrap at high resolution (Q Exactive or Fusion Lumos). Automatic gain control was used to control the number of ions accumulated in the respective ion traps. A single lock mass at *m/z* 445.120024 was used⁵³. All samples were analysed in technical duplicates.

Mass spectrometry data analysis

For the TAP-MS experiments (Fig. 1b, e), MS² spectra generated from +2, +3 or +4 charged precursor ions were extracted from the raw output files using msconvert from the ProteoWizard library (version 3.0.11220)⁵⁴. A protein search database was compiled on the basis of the UniProt *H. sapiens* reference proteome (release 2017_07; 71,591 entries)⁵⁵, extended by 248 typical contaminant sequences from MaxQuant (version 1.6.0.13)⁵⁶ and concatenated with reversed and shuffled decoy sequences generated by DecoyPyRat⁵⁷. Spectra were matched to semi-tryptic peptides of 6–40 amino acids of this database using MS-GF+ (version 2017.01.13)⁵⁸ allowing for a precursor mass error of up to 20 ppm and an isotope error of -1 to +2. All cysteines were considered carbamidomethylated and methionines optionally oxidated. Generated peptide-spectrum matches were post-processed based on the target-decoy approach⁵⁹ using Percolator (version 3.01)⁶⁰ and filtered for an estimated FDR of less than 1%. Protein groups implied by peptides across all samples were simplified using an in-house script based on the Occam's razor principle. The number of peptide-spectrum matches per gene (based on the UniProt annotation) and biological sample (collapsing technical replicates) served as input to SAINT (version 2)⁶¹ for probabilistic scoring of bait-prey interactions using the GFP samples as control. Significant interactions (FDR of less than 1%) were visualized as a network in Cytoscape 3.5.1⁶². The network was filtered to exclude non-specific interactors, showing an average spectral count of 10 or more in the 411 control samples of the CRAPome database (version 1.1)⁶³. For the immunoprecipitation-mass spectrometry experiments (Fig. 4g, Extended Data Fig. 9g), acquired raw data files were processed using the Proteome Discoverer 2.2.0.388 platform, using the database search engine Sequest HT. Percolator v.3.0 was used to remove false positives with an FDR of 1% on peptide and protein level under strict conditions. Searches were performed with full tryptic digestion against the human SwissProt database v.2017.06 (20,456 sequences and appended known contaminants) with up to two miscleavage sites. Oxidation (+15.9949 Da) of methionine was set as a variable modification, and carbamidomethylation (+57.0214 Da) of cysteine residues was set as a fixed modification. Data were searched with mass tolerances of ± 10 ppm and 0.025 Da on the precursor and fragment ions, respectively. Results were filtered to include peptide-spectrum matches with Sequest HT cross-correlation factor (Xcorr) scores of ≥ 1 and high peptide confidence. For Fig. 4g, obtained protein abundances by the Proteome Discoverer software were normalized for each sample to the corresponding mean bait abundance. On the basis of three biological replicate measurements, the following differences were tested using Welch's *t*-test (two-sided) on the normalized, log₂-transformed abundances, assuming normality but unequal variances. In the wild type, IRF5 abundance increases significantly upon CpG induction (effect size: +3.48; 95% confidence interval: [+3.03, +3.94]; *t*-statistic: 24.97; *P*-value: 2.09×10^{-4} ; degrees of freedom: 2.83). CpG-induced abundance of IRF5

Article

is significantly lower in the wild type compared to the point mutant (effect size: -2.20 ; 95% confidence interval: $[-2.57, -1.82]$; t -statistic: -18.16 ; P value: 2.77×10^{-4} ; degrees of freedom: 3.14).

Multiple sequence alignment and secondary structure prediction

Protein sequences were extracted from the UniProt⁶⁴ database and aligned with ClustalX2.1⁶⁵ using default settings. Secondary structure prediction was performed for the human TASL sequence using JPred4⁶⁶.

Gene expression analysis

THP1 control (sg*Ren*) and two knockout cell lines per gene (3×10^6 cells per point) were left untreated or stimulated for 6 h with $5 \mu\text{g/ml}$ R848 or $2 \mu\text{M}$ CpG-B (ODN2006). Cells were collected and RNA was isolated using the Qiagen RNeasy Mini kit including a DNase I digest step. RNA-sequencing (RNA-seq) libraries were prepared using QuantSeq 3' mRNA-Seq Library Prep Kit FWD for Illumina (Lexogen) according to the manufacturer's protocol. Libraries were subjected to 50-bp single-end high-throughput sequencing on an Illumina HiSeq 4000 platform at the Biomedical Sequencing Facility (<https://biomedical-sequencing.at/>). Raw sequencing reads were demultiplexed, and after barcode, adaptor and quality trimming with cutadapt (<https://cutadapt.readthedocs.io/en/stable/>), quality control was performed using FastQC (<http://www.bioinformatics.babraham.ac.uk/projects/fastqc/>). The remaining reads were mapped to the GRCh38/h38 human genome assembly using genomic short-read RNA-Seq aligner STAR version 2.5⁶⁷. We obtained more than 98% mapped reads in each sample with 70–80% of reads mapping to unique genomic location. Transcripts were quantified using End Sequence Analysis Toolkit (ESAT)⁶⁸. Differential expression analysis was performed using three biological replicates with DESeq2 (1.21.21) on the basis of read counts⁶⁹. Exploratory data analysis and visualizations were performed in R-project version 3.4.2 (Foundation for Statistical Computing, <https://www.R-project.org/>) with Rstudio IDE version 1.0.143, ggplot2 (3.0.0), dplyr (0.7.6), readr (1.1.1), gplots (3.0.1).

Transcription factor targets enrichment test

Tissue- and cell-type-specific high-level regulatory networks were extracted from the Network compendium (www.regulatorycircuits.org)⁷⁰. The network was filtered to extract strong transcription factor-to-target associations. Enrichment was analysed using Fisher's exact test and P values were corrected for multiple testing using an FDR procedure⁷¹. Background and gene sets are described in the corresponding figure legends.

GO enrichment analysis

GO biological process enrichment analysis for genes was performed using the Database for Annotation, Visualization and Integrated Discovery (DAVID) functional annotation tool, version 6.8⁷². Background and gene sets are described in the corresponding figure legends.

Expression database analysis of SLC15A4 and TASL

RNA-seq data for cancer cell lines were obtained from ref.²⁹ and grouped into tissues using the original annotation provided. Cap analysis gene expression (CAGE) data for primary cells were downloaded in April 2017 from the FANTOM5 website (<https://fantom.gsc.riken.jp/5/>)³⁰. Only transcript per million (TPM) values for p1 peaks were considered for each gene. Cell types were manually annotated to their corresponding tissues. Plotting was done using R.

Molecular docking

The TASL phosphomimetic pLxID-containing peptide (residues 286–299, ISTPSLHIDQYSNV) was docked on the structure of IRF5 (RCSB Protein Data Bank identifier (PDB ID) 3DSH) using CABS-dock⁷³. A constraint was applied to anchor the phosphomimetic residue to the

IRF5 residue R353 on the basis of the similarity between the IRF5 structure and IRF3 bound to the STING, MAVS or TRIF pLxIS-containing peptides¹¹. The top solution was further refined with the FlexPepDock algorithm⁷⁴ and the top ranking model compared to the structures of the IRF–peptide complexes.

Phagocytosis assay

Phagocytosis assays were carried out as previously described⁷⁵. Fluoresbrite carboxylated $1.75 \mu\text{M}$ microspheres (yellow green: 441-nm excitation, 486-nm emission, Polyscience, cat no. 17687-5) were opsonized in 50% human male AB serum in PBS for 16 h at 4°C under constant rotation. Beads were then washed twice with PBS and labelled with $2 \mu\text{g/ml}$ pHrodo-Red, SE (Thermo Fisher Scientific, cat. no. P36600) for 30 min at room temperature with agitation. Next, beads were washed once with PBS and were resuspended afterwards to a final concentration of 1×10^9 beads per ml.

THP1 cells were PMA-differentiated and seeded on 12-well cell-culture-coated dishes (1×10^6 cells per well). Labelled beads were then added at a ratio of 10 beads per cell and incubated for 3 h. Subsequently, cells were washed three times with ice-cold PBS and afterwards detached by scraping with a cell scraper (Sarstedt) and analysed by flow cytometry. For the Bafilomycin A1 (Enzo) control assay, the compound was added to the cell culture medium 30 min before addition of the labelled beads and added to the cell culture medium during the entire assay at a final concentration of 200 nM.

On flow cytometry, the intensity of the pH-insensitive dye (YG) and the intensity of pHrodo-Red, which reacts with an increase in signal to decreasing pH, were recorded. Cells that did not take up the labelled beads are negative for both signals and considered incapable of phagocytosis (PhagoNeg). Cells positive for the YG and a high pHrodo-Red signal underwent phagocytosis and phagosome acidification (PhagoLate). Cells positive only for the YG signal, and with a low pHrodo-Red signal, are in an early stage of phagocytosis (PhagoEarly).

Flow cytometry data acquisition was conducted on an LSR Fortessa II cytometer interfaced with FACSDiva (BD) and analysed using FlowJo software (v.10).

Quantification of LysoSensor intensities

For the imaging-based quantification of LysoSensor Green DND-189 (Thermo Fisher Scientific) in lysosomal compartments, 1×10^5 cells were co-stained with LysoSensor Green (1:1,000), LysoTracker Red DND-99 (1:10,000) (Thermo Fisher Scientific) and Hoechst 33342 (1:1,000) (Thermo Fisher Scientific) for 30 min in normal growth conditions. After pelleting by centrifugation, cells were resuspended in growth medium supplemented with $25 \mu\text{M}$ HEPES and transferred onto CellCarrier-384 Ultra Microplates (PerkinElmer). Following a brief centrifugation, cells were imaged on an Opera Phenix High-Content screening System (PerkinElmer) in confocal mode using the $63\times$ water immersion objective. Image analysis and lysosomal LysoSensor quantification were performed in CellProfiler version 3.1.5⁷⁶ and R version 3.4.4 (see Extended Data Fig. 6k for flow diagram). In brief, Hoechst 33342 and LysoSensor stainings were used for the detection of nuclei and cells. In a consecutive step, the LysoTracker signal allowed for the identification of lysosomes within cells and LysoSensor intensities were quantified within the identified lysosomal compartments. Plotting of the acquired data for visual representation was performed in R.

Reporting summary

Further information on research design is available in the Nature Research Reporting Summary linked to this paper.

Data availability

TAP-MS proteomics data have been deposited to the ProteomeXchange Consortium⁷⁷ via the PRIDE⁷⁸ partner repository with the dataset

identifier PXD014254 and 10.6019/PXD014254. RNA-seq data have been deposited to the Gene Expression Omnibus repository (GSE133317). Source data for immunoblots are provided in Supplementary Fig. 1.

77. Deutsch, E. W. et al. The ProteomeXchange consortium in 2017: supporting the cultural change in proteomics public data deposition. *Nucleic Acids Res.* **45**, D1100–D1106 (2017).
78. Perez-Riverol, Y. et al. The PRIDE database and related tools and resources in 2019: improving support for quantification data. *Nucleic Acids Res.* **47**, D442–D450 (2019).

Acknowledgements We thank all members of the G.S.-F. laboratory, T. Pernovska, G. Boehmelt, T. Krausgruber and E. Salzer for discussions and feedback; the Proteomics and Metabolomics Facility (CeMM) for the proteomics analyses; the Biomedical Sequencing Facility (CeMM/Medical University of Vienna) for the NGS sequencing; the Core Facility Imaging of the Medical University of Vienna; and T. Maeda for kindly providing the CAL-1 cells. This work was supported by Boehringer Ingelheim (Research Collaboration Agreement BI-CeMM 238114 to L.X.H., UK, K.P., A.S. and M.R.), the Austrian Academy of Sciences (to G.S.-F., F.K., V.S. and A.C.-R.), the European Research Council (ERC AdG 695214 GameofGates to G.S.-F., M.R., S.S., P.E., U.G., M.D.P., A.B. and E.G.) and Austrian Science Fund (FWF SFB F4711,

to J.W.B.). Plasmids obtained through Addgene were a gift from F. Zhang, D. Trono and D. Golenbock.

Author contributions L.X.H., J. Lee, E.G., M.L.M., C.E.W., M.R. and G.S.-F. designed research; L.X.H., J. Lee, U.K., K.P., F.K., P.E., A.S., S.S., A.C.M., F.J.K., M.D.P., E.G. and M.R. performed research; V.S., A.C.-R. and U.G. performed bioinformatic analysis; J.W.B., A.B. and J. Li generated reagents and provided scientific insight; L.X.H., M.R. and G.S.-F. analysed and interpreted the data; L.X.H., M.R. and G.S.-F. wrote the paper.

Competing interests J. Lee, J. Li, J.K., M.L.M. and C.E.W. are employed in the Immunology and Respiratory department, Drug Concept Discovery group of Boehringer-Ingelheim. The work in the G.S.-F. laboratory is supported by Boehringer-Ingelheim (grant agreement 238114).

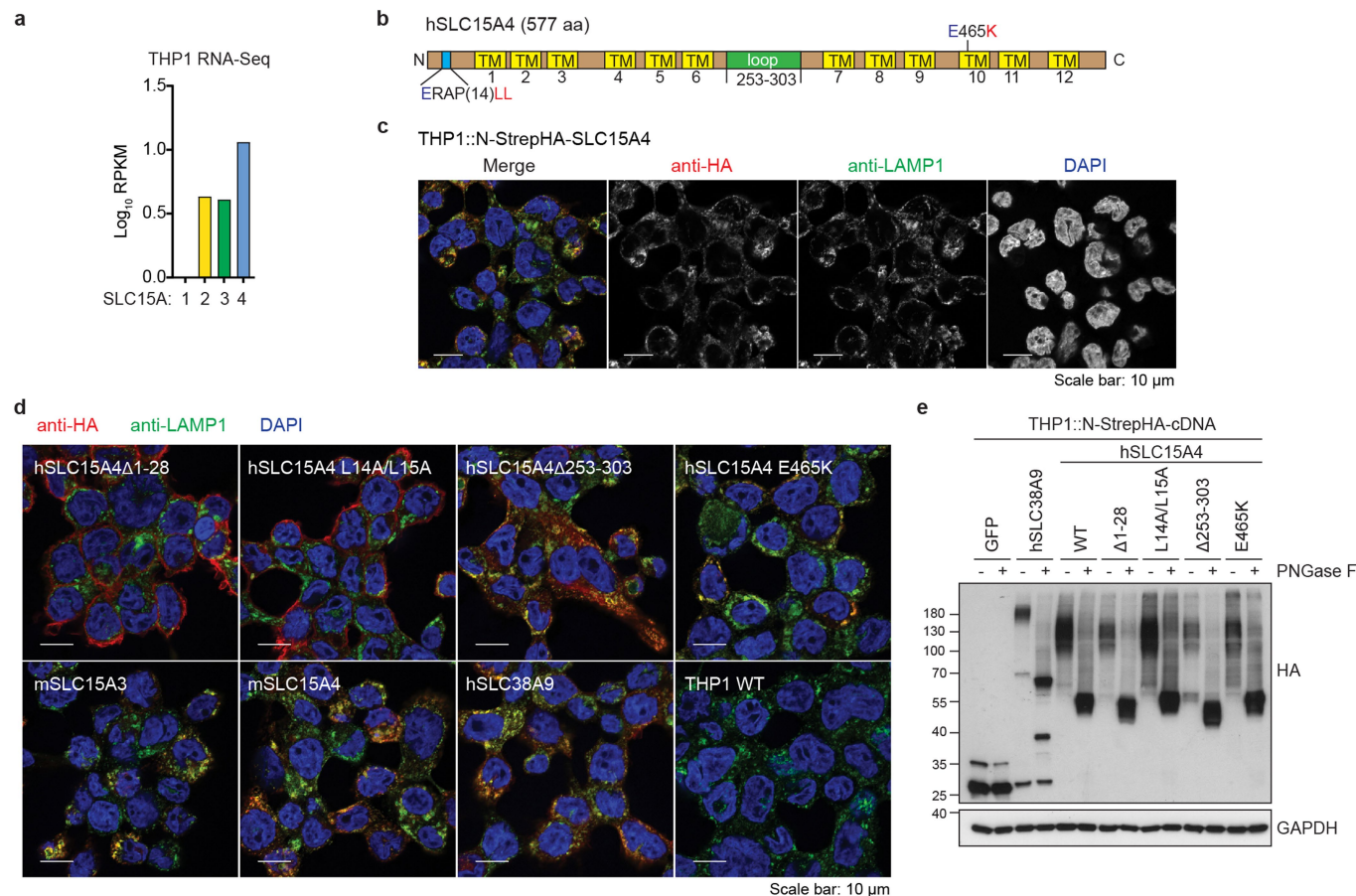
Additional information

Supplementary information is available for this paper at <https://doi.org/10.1038/s41586-020-2282-0>.

Correspondence and requests for materials should be addressed to M.R. or G.S.-F.

Peer review information *Nature* thanks Zhijian (James) Chen and the other, anonymous, reviewer(s) for their contribution to the peer review of this work.

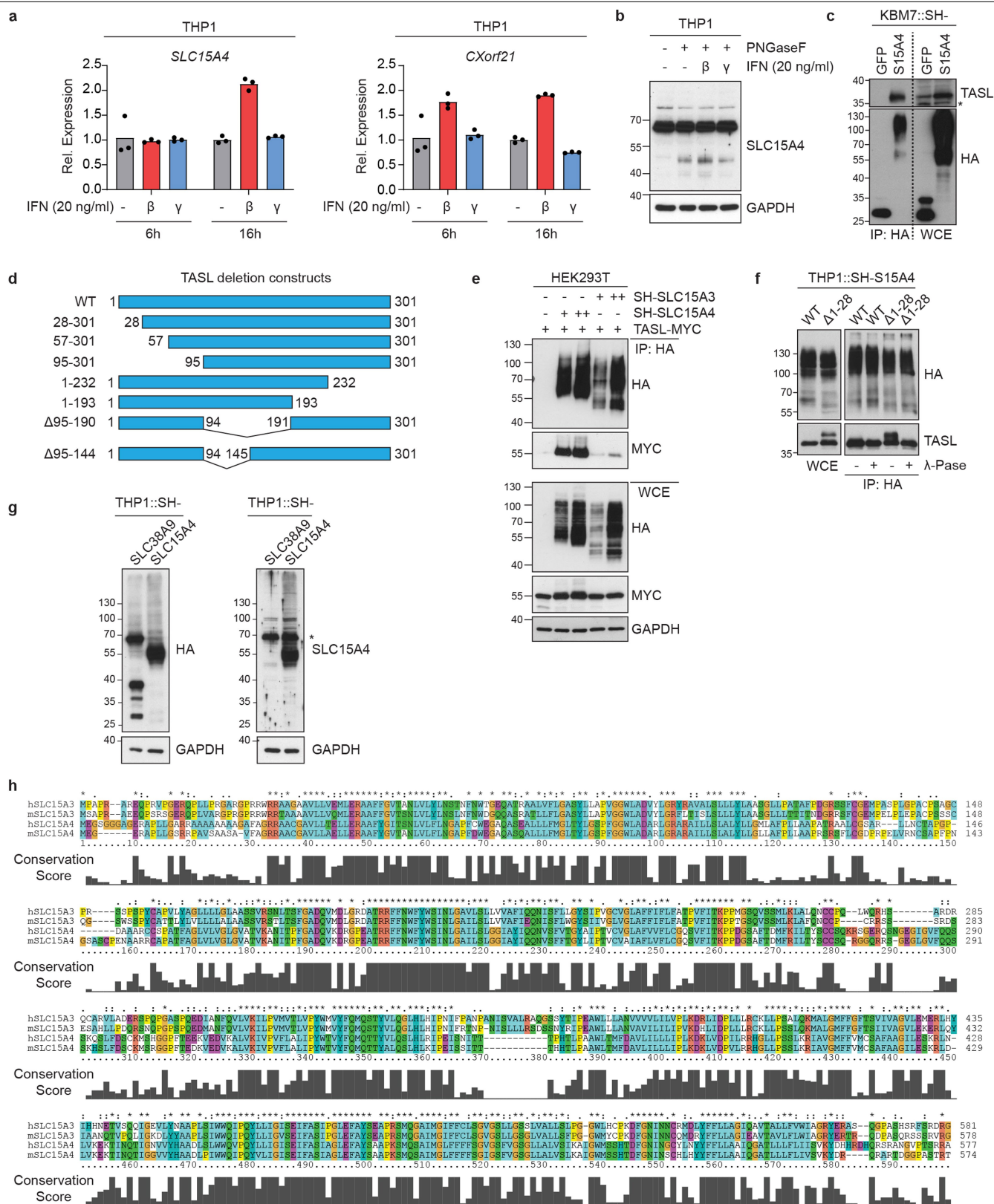
Reprints and permissions information is available at <http://www.nature.com/reprints>.



Extended Data Fig. 1 | Subcellular localization of tagged SLC15A4 and related constructs in THP1 cells. **a**, Gene expression levels of SLC15A1, SLC15A2, SLC15A3 and SLC15A4 in THP1 cells from a previous publication²⁹. **b**, Domain organization of SLC15A4 protein. TM, transmembrane domain. **c**, **d**, Confocal microscopy of indicated THP1 cells. Red, anti-HA; green,

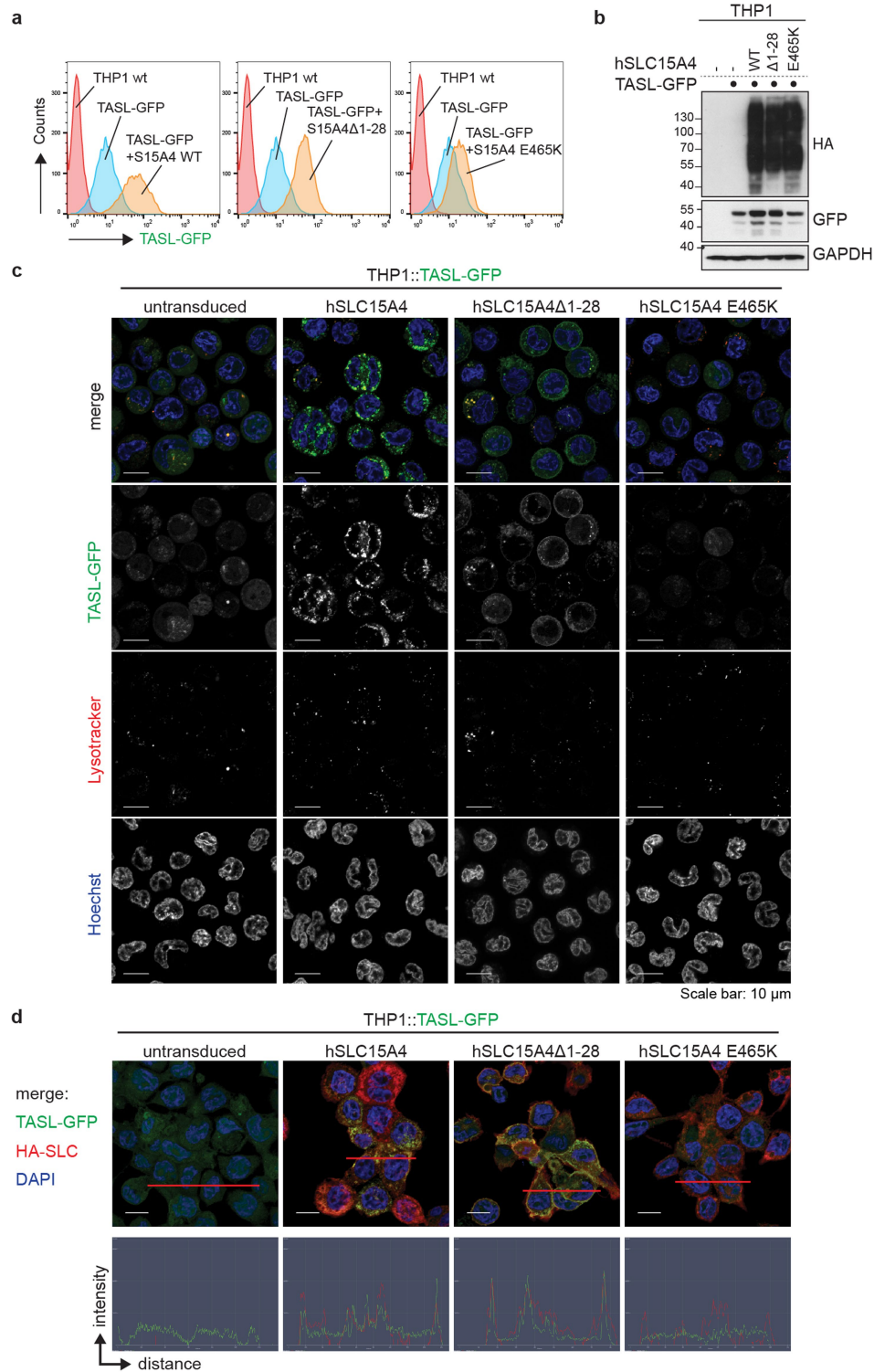
anti-LAMP1; blue, DAPI. Scale bars, 10 μm. **e**, Lysates from indicated THP1 cells untreated or treated with PNGase F were analysed by immunoblotting. **c–e**, Data representative of two independent experiments. For gel source data, see Supplementary Fig. 1.

levels of *SLC15A4* and *CXorf21* in primary human cells as measured by CAGE by FANTOM³⁰ **(b)** and in human cancer cell lines measured by RNA-seq from a previous publication²⁹ **(c)**. Circles represent individual samples; in box plots, bars indicate the median, boxes indicate the first to third quartiles. The top whisker extends from hinge to largest value no further than 1.5× interquartile range (IQR) from the hinge, and the bottom whisker extends from the hinge to smallest value at most 1.5× IQR of the hinge.



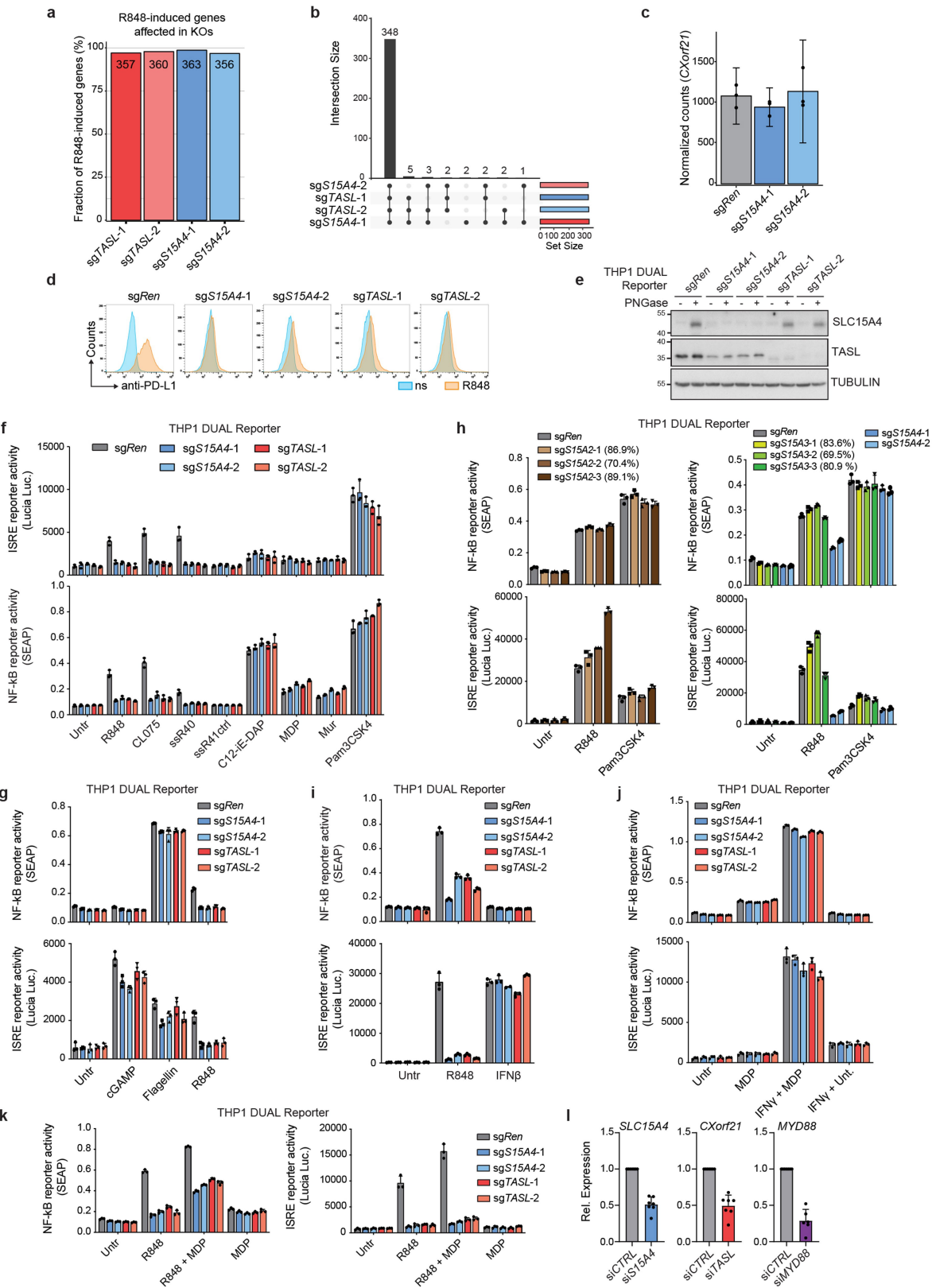
Extended Data Fig. 3 | Type-I-interferon-inducible TASL forms a complex with SLC15A4. a, Normalized mRNA expression of *SLC15A4* and *CXorf21* relative to *GAPDH* in THP1 cells treated as indicated. Data show mean ($n=3$ technical replicates). **b**, Immunoblots of lysates of THP1 cells stimulated with interferon- β or interferon- γ (20 ng ml $^{-1}$, for 16 h) treated with PNGase F as indicated. **c, e**, Lysates from KBM7 cells transduced (c) or HEK293T cells transiently transfected (e), as indicated, were subjected to HA immunoprecipitation. Immunoprecipitates and whole-cell extracts were

analysed by immunoblotting. **d**, Overview of deletion mutants used in Fig. 1g. **f**, Lysates from indicated THP1 cells were subjected to HA immunoprecipitation and treated or not with λ phosphatase. Immunoprecipitates and whole-cell extract were analysed by immunoblotting. **g**, Immunoblots of indicated THP1 cells treated with PNGase F. **h**, Multiple sequence alignment of human and mouse SLC15A3 and SLC15A4. UniProt entry names: S15A3_HUMAN, S15A3_MOUSE, S15A4_HUMAN, S15A4_MOUSE. In a–c, e–g, data are representative of two independent experiments. For gel source data, see Supplementary Fig. 1.



Extended Data Fig. 4 | SLC15A4 controls TASL protein abundance and subcellular localization. a, b, Flow cytometry (a) and immunoblot (b) of indicated THP1 cells. **c,** Confocal live microscopy of indicated THP1 cells. Green, TASL-GFP; red, Lysotracker; blue, Hoechst33342. Scale bars, 10 μ m. **d,** Top, confocal microscopy of indicated formaldehyde-fixed THP1 cells.

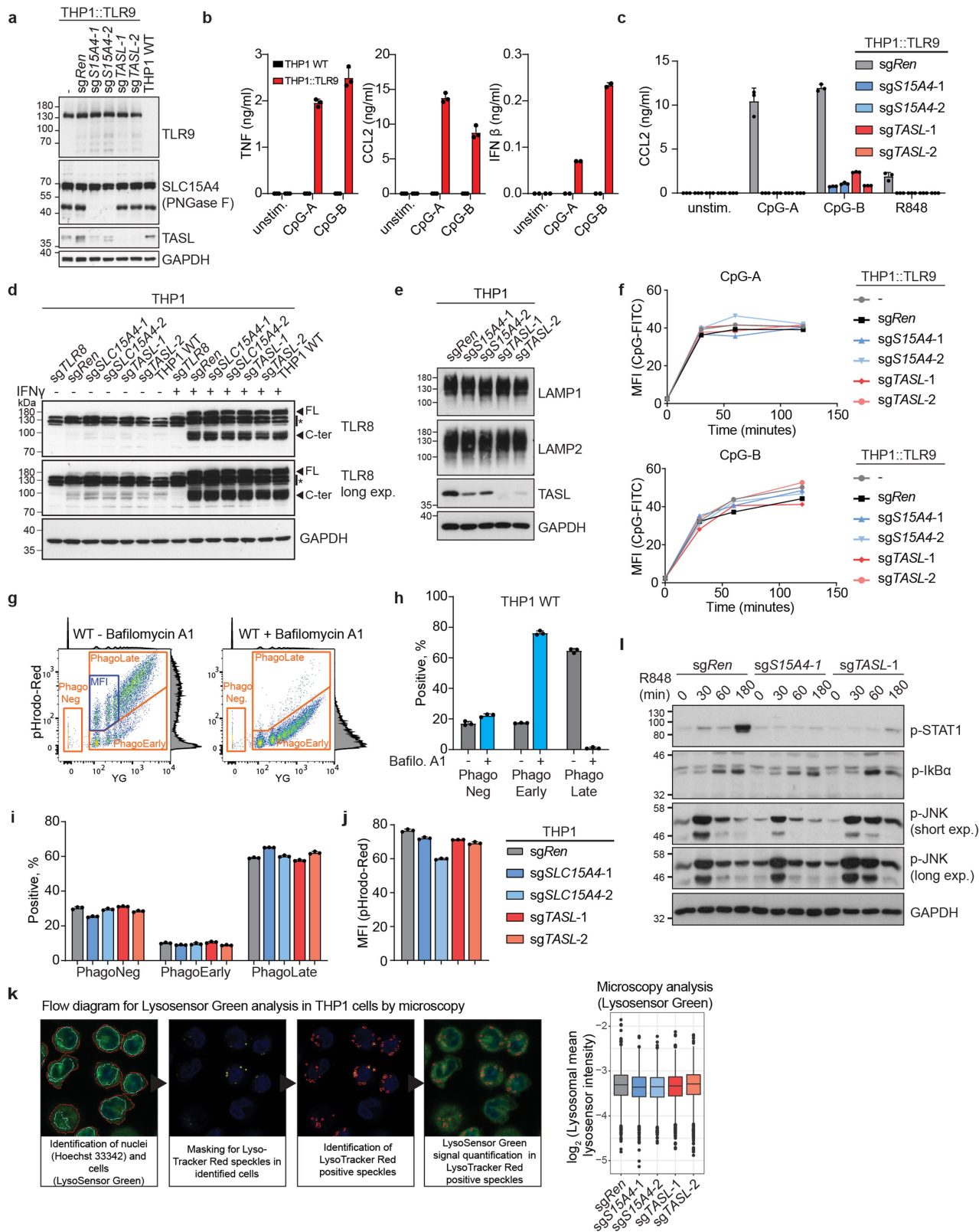
Green, TASL-GFP; red, HA; blue, DAPI. Scale bars, 10 μ m. Bottom, profiles of signal intensity of TASL-GFP (green) and HA (red) along the red lines shown in microscopy images (top). In a–d, data are representative of two independent experiments. For gel source data, see Supplementary Fig. 1.



Extended Data Fig. 5 | See next page for caption.

Extended Data Fig. 5 | TASL mirrors SLC15A4 requirement for TLR7 and TLR8 activation. **a**, Fraction of R848-induced genes affected by *SLC15A4* and *CXorf21* knockout, related to Fig. 2b. **b**, Upset plot representing number of R848-induced genes commonly affected by the indicated sgRNAs, related to Fig. 2b. **c**, *CXorf21* gene expression levels in indicated THP1 cells, related to Fig. 2b. Bar graphs shown mean ($n = 3$ biological replicates), error bars show 95% confidence interval of mean. **d**, Flow cytometry of PD-L1 surface expression in indicated unstimulated (ns) or R848-stimulated ($5 \mu\text{g ml}^{-1}$, for 24 h) THP1 cells. **e**, Immunoblots of indicated THP1 DUAL cells. Lysates treated with PNGase F, as indicated. **f–i, k**, Indicated THP1 DUAL cells were (co-)treated for 24 h with R848 ($5 \mu\text{g ml}^{-1}$), CL075 ($5 \mu\text{g ml}^{-1}$), single-stranded (ss)RNA40 complexed with LyoVec ($5 \mu\text{g ml}^{-1}$) or inactive control ssRNA41 with LyoVec

($5 \mu\text{g ml}^{-1}$), C12-iE-DAP ($5 \mu\text{g ml}^{-1}$), MDP ($10 \mu\text{g ml}^{-1}$), murabutide ($10 \mu\text{g ml}^{-1}$), Pam3CSK4 ($0.1 \mu\text{g ml}^{-1}$), flagellin ($0.1 \mu\text{g ml}^{-1}$), cGAMP ($3 \mu\text{g ml}^{-1}$) or interferon- β (20 ng ml^{-1}). **h**, CRISPR-Cas9 editing efficiency (%) estimated by TIDE. **j**, Indicated THP1 DUAL cells were primed or not with interferon- γ ($0.1 \mu\text{g ml}^{-1}$) for 24 h, washed and stimulated or not with MDP ($10 \mu\text{g ml}^{-1}$, for 24 h). **f–k**, Supernatants were analysed for ISRE and NF- κB reporter activity. Mean \pm s.d. ($n = 3$ biological replicates). **i**, Relative mRNA expression of *SLC15A4*, *CXorf21* or *MYD88* in siRNA-transfected CD14⁺ monocytes in comparison to control (siCTRL). Data represent mean \pm s.d. from six (*MYD88*) or seven (*SLC15A4* and *CXorf21*) individual donors. In **d–k**, data are representative of two independent experiments. For gel source data, see Supplementary Fig. 1.

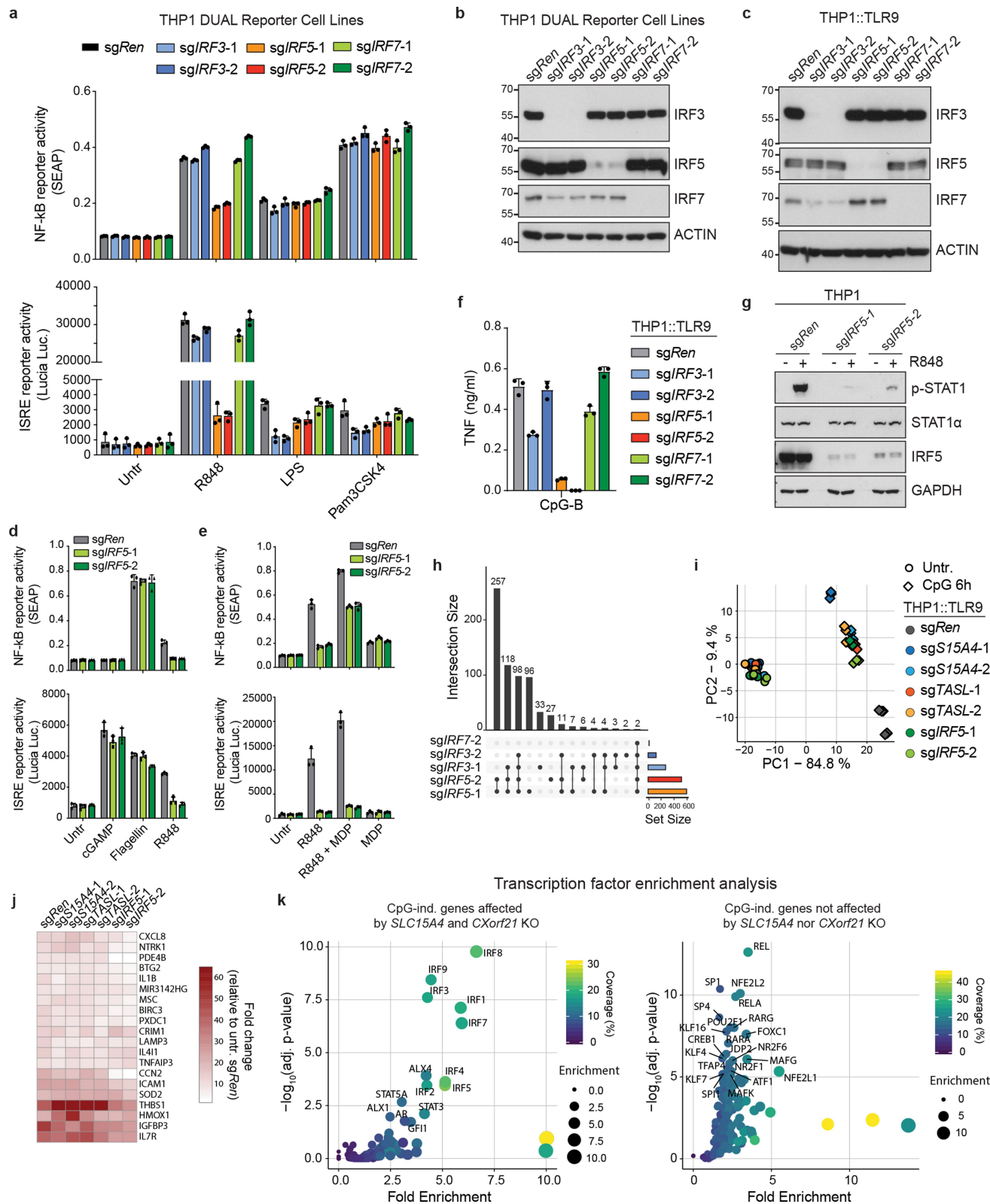


Extended Data Fig. 6 | See next page for caption.

Extended Data Fig. 6 | TASI and SLC15A4 deficiency impairs endosomal TLR-mediated signalling downstream of receptor engagement.

a, Immunoblots of lysates of THP1 cells treated with PNGase F, as indicated. **b, c**, Cytokine production of indicated THP1 cells unstimulated or stimulated with CpG-A, CpG-B (5 μ M) or R848 (5 μ g ml⁻¹) for 24 h. Data show mean \pm s.d. of biological replicates (TNF and CCL2, $n=3$; IFN β , $n=2$). **d**, Immunoblots of indicated THP1 cells stimulated or not with interferon- γ (0.1 μ g ml⁻¹, for 16 h). **e**, Immunoblots of indicated THP1 cells. **f**, Indicated THP1::TLR9 cells treated with FITC-labelled CpG-A or CpG-B (1 μ M, for 0–120 min) were analysed by flow cytometry. **g**, Representative flow cytometry scatter plots of phagocytosis assays. Differentiated THP1 cells, treated or not with bafilomycin A1, were incubated with dual-coloured opsonized beads. Using intensities of pH-insensitive (YG) and pH-sensitive (pHrodo-Red, signal increases with decrease in pH) dyes, cells are divided into phagocytosis-negative (PhagoNeg, double-negative), cells that have undergone phagocytosis and phagosome acidification (PhagoLate, double-positive) and early phagocytic cells (PhagoEarly, YG and low pHrodo-Red signal). The marginal intensity distributions are displayed on the sides of the plot. **h**, Bar graphs show

mean \pm s.d. ($n=3$ biological replicates) of fractions described in **g**. **i, j**, Indicated THP1 cells were subjected to phagocytosis assays. **i**, Bar graphs show mean \pm s.d. ($n=3$ biological replicates) of fractions described in **g**. **j**, Bar graphs represent mean \pm s.d. ($n=3$ biological replicates) of the mean fluorescence intensities (MFI) of the pHrodo-Red signal acquired in the MFI gate shown in **g**, focusing on cells having taken up 1–3 beads per cell. **k**, Flow diagram for quantification of Lysosensor Green intensities in lysosomal compartments by microscopy. Box plots show intensity of Lysosensor signal on Lysotracker-positive lysosomes, as measured by imaging in the indicated THP1 cells. Bars indicate median, boxes indicate the first to third quartiles; the top whisker extends from the hinge to the largest value no further than 1.5 \times IQR from the hinge; the bottom whisker extends from the hinge to the smallest value at most 1.5 \times IQR of the hinge. Outliers are shown as circles. *sgRen*, $n=2,432$; *sgSLC15A4-1*, $n=1,721$; *sgSLC15A4-2*, $n=1,981$; *sgTASL-1*, $n=2,378$; *sgTASL-2*, $n=2,627$ quantified speckles. **l**, Immunoblots of indicated THP1 cells stimulated with R848 (5 μ g ml⁻¹, for 0–180 min.). In **a–k**, data are representative of two (**a–f, k, l**) or three (**g–j**) independent experiments. For gel source data, see Supplementary Fig. 1.



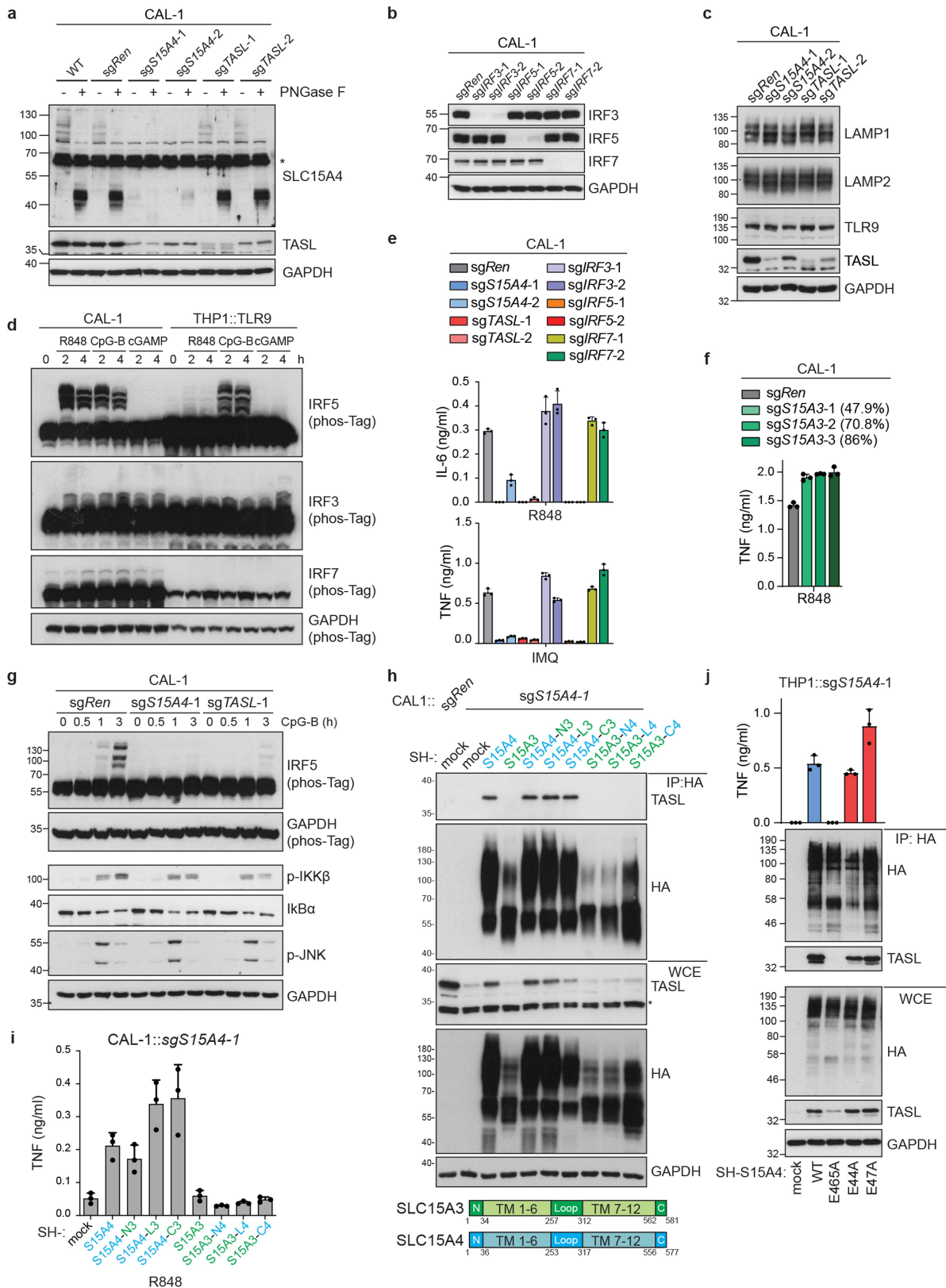
Extended Data Fig. 7 | See next page for caption.

Extended Data Fig. 7 | Loss of TASL or SLC15A4 mirrors IRF5 deficiency in perturbing endosomal TLR responses. a, d, e,

THP1 DUAL cells were (co-) treated for 24 h with R848 ($5 \mu\text{g ml}^{-1}$), LPS ($0.1 \mu\text{g ml}^{-1}$), Pam3CSK4 ($0.1 \mu\text{g ml}^{-1}$), cGAMP ($3 \mu\text{g ml}^{-1}$), flagellin ($0.1 \mu\text{g ml}^{-1}$) or MDP ($10 \mu\text{g ml}^{-1}$) as indicated. Supernatants were analysed for ISRE and NF- κ B reporter activity.

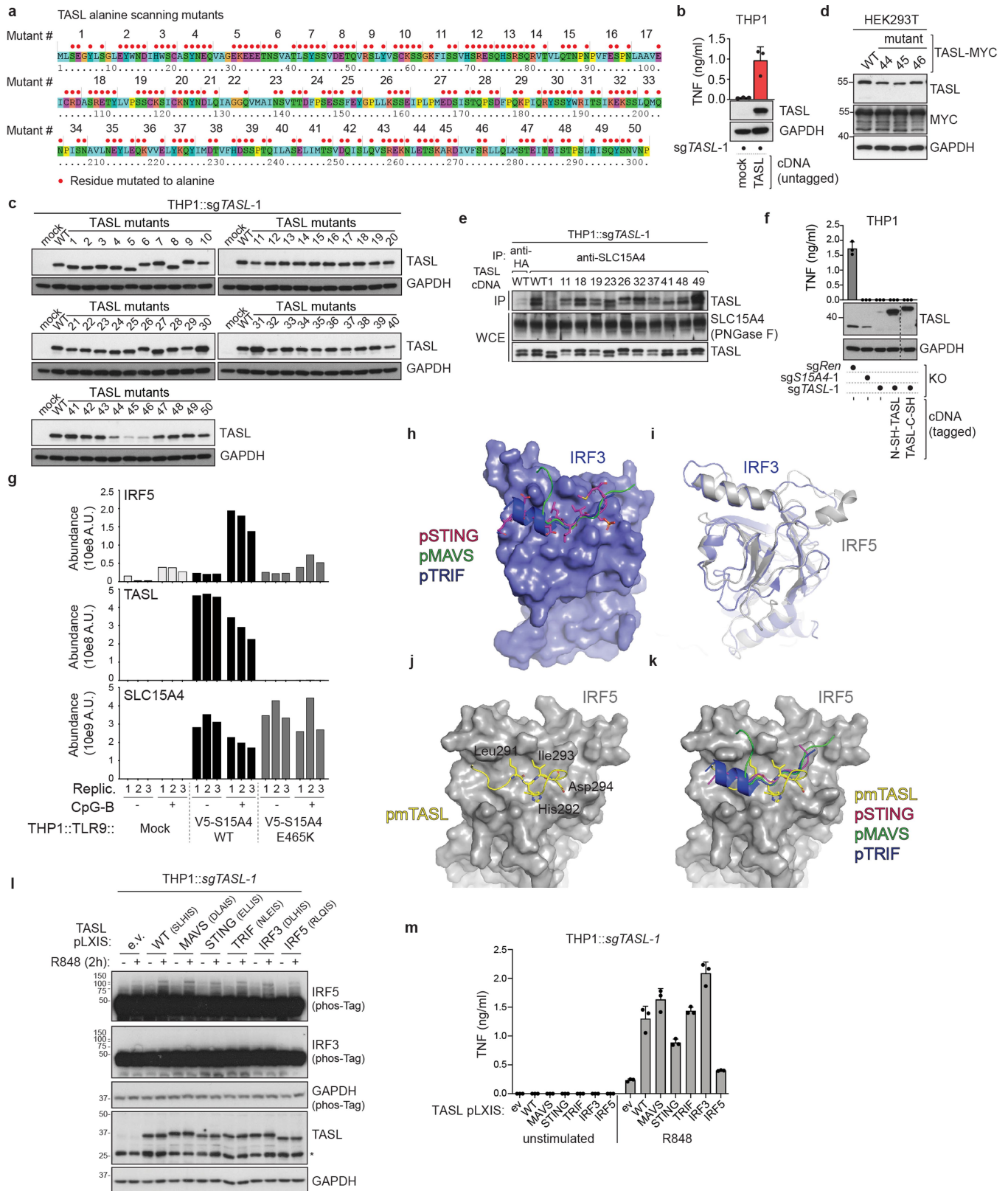
b, c, Immunoblots of indicated THP1 DUAL (**b**) or THP1::TLR9 (**c**) cells. **f,** TNF production of indicated THP1::TLR9 cells stimulated with CpG-B ($2 \mu\text{M}$, for 24 h). **g,** Immunoblots of indicated THP1 cells stimulated or not with R848 ($5 \mu\text{g ml}^{-1}$, for 3 h). **h,** Upset plot representing number of CpG-B-induced genes ($2 \mu\text{M}$, for 6 h) (DESeq2 adjusted P value < 0.05 , $n = 3$ biological replicates) in comparison to control (*sgRen*) commonly affected by indicated sgRNAs. No gene was significantly affected by *sgIRF7-1*. **i,** Principal component analysis plot of transcriptional profiles of untreated and CpG-B-treated ($2 \mu\text{M}$, for 6 h)

THP1::TLR9 cells ($n = 3$ biological replicates) shown in Fig. 3d. **j,** Heat map representing 20 most-induced genes by CpG-B in control THP1::TLR9 cells and not affected by *SLC15A4*, *CXorf21* or *IRF5* knockout, related to Fig. 3d, e. **k,** Transcription factor enrichment analysis (two-sided Fisher's exact test, P value adjusted for multiple testing) of genes upregulated upon CpG-B treatment in control THP1 cells specifically affected (DESeq2 adjusted P value < 0.05 , $n = 3$ biological replicates) (left) or not (right) by *SLC15A4* and *CXorf21* knockout, related to Fig. 3d, e. Background sets are defined as all genes upregulated by CpG-B treatment or all expressed genes (counts per million > 1) respectively. In **a, d–f**, data are mean \pm s.d. ($n = 3$ biological replicates). In **a–g**, data are representative of two independent experiments. For gel source data, see Supplementary Fig. 1.



Extended Data Fig. 8 | The SLC15A4–TASL complex is required for IRF5-dependent signalling in human CAL-1 plasmacytoid dendritic cells. **a–c**, Immunoblots of lysates of CAL-1 cells, with PNGase F treatment as indicated. **d**, Immunoblots of CAL-1 or THP1::TLR9 cells stimulated with R848 ($5 \mu\text{g ml}^{-1}$), CpG-B ($5 \mu\text{M}$) or cGAMP ($3 \mu\text{g ml}^{-1}$) as indicated. **e, f**, Cytokine production of indicated CAL-1 cells stimulated with R848 ($5 \mu\text{g ml}^{-1}$) or imiquimod (IMQ) ($5 \mu\text{g ml}^{-1}$) for 24 h. **f**, CRISPR–Cas9 editing efficiency (%) estimated by TIDE. **g**, Immunoblots of CAL-1 cells stimulated with CpG-B ($5 \mu\text{M}$,

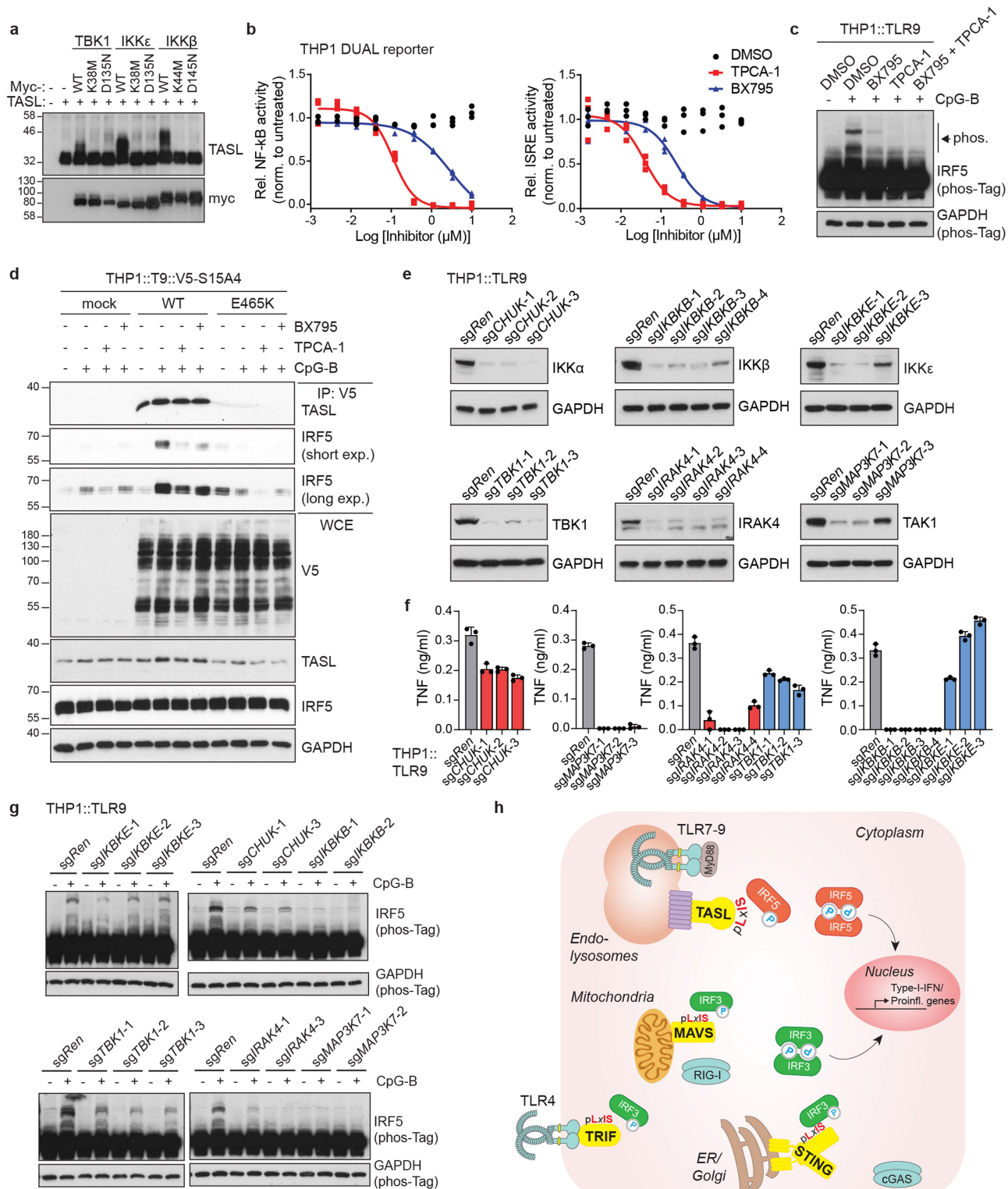
for 0–180 min) as indicated. **h**, Lysates from indicated CAL-1 cells were subjected to HA immunoprecipitation and analysed by immunoblotting. **i**, TNF production of cells described in **h** upon stimulation with R848 ($5 \mu\text{g ml}^{-1}$, for 24 h). **j**, Immunoblots show expression levels (whole-cell extracts) and HA immunoprecipitates from indicated THP1 cells. Bar graphs represent TNF production upon R848 stimulation ($5 \mu\text{g ml}^{-1}$, for 24 h). In **e, f, i, j**, bar graphs show mean \pm s.d. ($n = 3$ biological replicates). In **a–j**, data are representative of two independent experiments. For gel source data, see Supplementary Fig. 1.



Extended Data Fig. 9 | See next page for caption.

Extended Data Fig. 9 | Mutagenesis of TASL identifies functional elements and reveals a pLxIS motif required for IRF5 activation. **a**, Overview of mutants used in Fig. 4b; changes to alanine indicated by red circles. **b, f**, Immunoblots of indicated THP1 cells. Bar graphs represent TNF levels following R848 stimulation ($5 \mu\text{g ml}^{-1}$, for 24 h). In **f**, the dashed line indicates cropping of unrelated lanes from the same blots. **c**, Immunoblots of indicated reconstituted TASL-deficient THP1 cells. **d**, Normal expression levels, but reduced detection by anti-TASL antibodies, of TASL mutants targeting amino acids 261–277. Immunoblots of HEK293T cells transiently transfected with indicated cDNAs. **e**, Lysates from indicated THP1 cells were subjected to immunoprecipitation. Immunoprecipitates and whole-cell extracts were analysed by immunoblotting. **g**, Abundance of indicated proteins determined by mass spectrometry in V5 immunoprecipitates from THP1::TLR9 cells stimulated with CpG-B ($5 \mu\text{M}$, for 2 h) as indicated, related to Fig. 4g. Three

biological replicates are shown. **h**, Crystal structures of IRF3 bound to phosphorylated pLxIS-containing peptides from STING (pink, PDB ID: 5JEJ), MAVS (green, PDB ID: 5JEK) and TRIF (blue, PDB ID: 5JEL). Residues in STING peptides are shown as sticks. **i**, Superposition of peptide-bound IRF3 (PDB ID: 5JEJ) and dimeric IRF5 (one monomer shown, PDB ID: 3DSH), showing highly similar folds. **j**, Model of phosphomimetic pLxID-containing peptide from TASL (pmTASL, ISTPSLHIDQYSNV, yellow) bound to IRF5. Residues corresponding to the pLxID motif shown as sticks. **k**, Comparison of binding mode of pLxIS-containing peptides to IRF proteins. Only IRF5 is shown for clarity. **l**, Immunoblots of indicated THP1 cells unstimulated or stimulated with R848 ($5 \mu\text{g ml}^{-1}$, for 2 h). **m**, TNF production of cells described in **l**, stimulated with R848 ($5 \mu\text{g ml}^{-1}$, for 24 h). In **b, f, l, m**, bar graphs show mean \pm s.d. ($n = 3$ biological replicates). In **b–f, l, m**, data are representative of two independent experiments. For gel source data, see Supplementary Fig. 1.



Reporting Summary

Nature Research wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Research policies, see [Authors & Referees](#) and the [Editorial Policy Checklist](#).

Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

- | | |
|-------------------------------------|--|
| n/a | Confirmed |
| <input type="checkbox"/> | <input checked="" type="checkbox"/> The exact sample size (n) for each experimental group/condition, given as a discrete number and unit of measurement |
| <input type="checkbox"/> | <input checked="" type="checkbox"/> A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly |
| <input type="checkbox"/> | <input checked="" type="checkbox"/> The statistical test(s) used AND whether they are one- or two-sided
<i>Only common tests should be described solely by name; describe more complex techniques in the Methods section.</i> |
| <input type="checkbox"/> | <input checked="" type="checkbox"/> A description of all covariates tested |
| <input type="checkbox"/> | <input checked="" type="checkbox"/> A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons |
| <input type="checkbox"/> | <input checked="" type="checkbox"/> A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals) |
| <input type="checkbox"/> | <input checked="" type="checkbox"/> For null hypothesis testing, the test statistic (e.g. F , t , r) with confidence intervals, effect sizes, degrees of freedom and P value noted
<i>Give P values as exact values whenever suitable.</i> |
| <input checked="" type="checkbox"/> | <input type="checkbox"/> For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings |
| <input checked="" type="checkbox"/> | <input type="checkbox"/> For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes |
| <input checked="" type="checkbox"/> | <input type="checkbox"/> Estimates of effect sizes (e.g. Cohen's d , Pearson's r), indicating how they were calculated |

Our web collection on [statistics for biologists](#) contains articles on many of the points above.

Software and code

Policy information about [availability of computer code](#)

Data collection	ELISA and Luciferase/SEAP reporter data: Molecular Devices, SoftMax Pro v7.0 Flow cytometry: BD CellQuest Pro v6.0; BD FACS DIVA v8.0.1 RealTime PCR: QIAGEN, Rotor Gene Series Software Version 2.2.2 and Roche LightCycler Software v1.5 Microscopy: Carl Zeiss AG, ZEN 2.3; PerkinElmer Harmony 4.9 Proteomics: Thermo Fisher Scientific, Xcalibur v2.1.0 SP1, v4.1.31.9 and v4.2.28.14
Data analysis	Microscopy: Carl Zeiss AG, ZEN 2.3, CellProfiler version 3.1.5 Flow cytometry: FlowJo v10 Multiple sequence alignments: ClustalX2 qPCR: QIAGEN, Rotor Gene Series Software Version 2.2.2 and Roche LightCycler Software v1.5; calculations were performed in Excel 2016 Data graphs for ELISA, qPCR, Luciferase and SEAP reporter assays were prepared in GraphPad Prism v7 or Excel 2016 Proteomics: Thermo Fisher Scientific, Proteome Discoverer 2.2.0.388 platform

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors/reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Research [guidelines for submitting code & software](#) for further information.

Data

Policy information about [availability of data](#)

All manuscripts must include a [data availability statement](#). This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A list of figures that have associated raw data
- A description of any restrictions on data availability

TAP-MS proteomics data have been deposited to the ProteomeXchange Consortium via the PRIDE partner repository with the dataset identifier PXD014254 and

Field-specific reporting

Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

- ☒ Life sciences
- ☐ Behavioural & social sciences
- ☐ Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see [nature.com/documents/nr-reporting-summary-flat.pdf](https://www.nature.com/documents/nr-reporting-summary-flat.pdf)

Life sciences study design

All studies must disclose on these points even when the disclosure is negative.

Sample size	Sample size was determined empirically, based on exploratory experiments, our previously published work as well as published literature with similar methodology. The sample sizes were considered sufficient due to the large effect sizes which allowed the biological interpretation of the results obtained.
Data exclusions	No data were excluded from the study.
Replication	Reported experiments were repeated at least 2 times with comparable results.
Randomization	Samples were not randomized for this study since no suggestive rating of data was involved.
Blinding	Blinding was not relevant for this study since no suggestive rating of data was involved.

Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

Materials & experimental systems	Methods
n/a	Involvement in the study
<input type="checkbox"/> <input checked="" type="checkbox"/> Antibodies	<input checked="" type="checkbox"/> <input type="checkbox"/> ChIP-seq
<input type="checkbox"/> <input checked="" type="checkbox"/> Eukaryotic cell lines	<input type="checkbox"/> <input checked="" type="checkbox"/> Flow cytometry
<input checked="" type="checkbox"/> <input type="checkbox"/> Palaeontology	<input checked="" type="checkbox"/> <input type="checkbox"/> MRI-based neuroimaging
<input checked="" type="checkbox"/> <input type="checkbox"/> Animals and other organisms	
<input checked="" type="checkbox"/> <input type="checkbox"/> Human research participants	
<input checked="" type="checkbox"/> <input type="checkbox"/> Clinical data	

Antibodies

Antibodies used	<p>Antibodies used in western blot</p> <p>Rabbit anti-HA (Cell Signaling, Cat.#: 3724, Lot #: 9, Dilution: 1:1000), rabbit anti-V5 (Cell Signaling, Cat.#: 13202), Lot #: 4, Dilution 1:1000), rabbit anti-GFP (Cell Signaling, Cat.#: 2956, Lot #: 4, Dilution 1:1000), rabbit anti-RAGA (Cell Signaling, Cat.#: 4357, Lot #: 2, Dilution 1:1000), rabbit anti-TLR9 (Cell Signaling, Cat.#: 5845, Lot #: 1, Dilution 1:5000), rabbit anti-TAK1 (Cell Signaling, Cat.#: 4505, Lot #: 4, Dilution 1:1000), rabbit anti-IRAK4 (Cell Signaling, Cat.#: 4363, Lot #: 4, Dilution 1:1000), rabbit anti-IKKα (Cell Signaling, Cat.#: 2682, Lot #: 5, Dilution 1:1000), rabbit anti-IKKβ (Cell Signaling, Cat.#: 8943, Lot #: 4, Dilution 1:1000), rabbit anti-TBK1 (Cell Signaling, Cat.#: 51872, Lot #: 4, Dilution 1:1000), rabbit anti-IKKe (Cell Signaling, Cat.#: 2905, Lot #: 3, Dilution 1:1000), rabbit anti-IRF3 (Cell Signaling, Cat.#: 11904, Lot #: 3, Dilution 1:1000), rabbit anti-IRF7 (Cell Signaling, Cat.#: 4920, Lot #: 2, Dilution 1:1000), rabbit anti-phospho-IKKα/β Ser176/180 (Cell Signaling, Cat.#: 2697, Lot #: 19, Dilution 1:1000), rabbit anti-phospho-IkBa Ser32/36 (Cell Signaling, Cat.#: 9246, Lot #: 16, Dilution 1:1000), rabbit anti-phospho-NF-κB p65 Ser536 (Cell Signaling, Cat.#: 3033, Lot #: 12, Dilution 1:1000), rabbit anti-phospho-p38 MAPK Thr180/Tyr182 (Cell Signaling, Cat.#: 9211, Lot #: 19, Dilution 1:1000), rabbit anti-phospho-SAPK/JNK Thr183/Tyr185 (Cell Signaling, Cat.#: 4668, Lot #: 15, Dilution 1:1000) Rabbit phospho-STAT1 Y701 (Cell Signaling, Cat.#: 7649, Lot #: 5, Dilution 1:1000), rabbit anti-CXorf21/TASL (Sigma, Cat.#: HPA001185, Lot #: A106306, Dilution 1:1000), rabbit anti-myc (Sigma, Cat.#: C3956, Lot #: 098k4806, Dilution 1:2000), rabbit anti-TLR8 (MBL, Cat.#: pd047, Lot #: 001, Dilution 1:1000), mouse anti-GAPDH (Santa Cruz, Cat.#: sc-365062, Lot #: H0515, Dilution 1:2000), rabbit anti-IkB-α (Santa Cruz, Cat.#: sc-371, Lot #: B1215, Dilution 1:1000), mouse anti-LAMP2 (Santa Cruz, Cat.#: sc-18822, Lot #: I1212, Dilution 1:1000), mouse anti-STAT1α (Santa Cruz, Cat.#: sc-417, Lot #: B0304, Dilution 1:1000), rabbit anti-Actin Cytoskeleton, Cat.#: AAN01-A, Lot #: 121, Dilution 1:2000), mouse anti-Tubulin (Abcam, Cat.#: ab7291, Lot #: GR3281114-3, Dilution 1:5000), rabbit anti-IRF5 (Abcam, Cat.#: ab181553, Lot #: GR3248905-3, Dilution 1:1000), mouse anti-Lamp1 (Abcam, Cat.#: ab25630, Lot #: 792614, Dilution 1:1000), custom rabbit anti-SLC15A4 raised against the N-terminus (GenScript).</p>
-----------------	--

Antibodies used in immunoprecipitation

Rabbit anti-SLC38A9 (Sigma, Cat.#: HPA043785, Lot #: A96351, 1:250), rabbit anti-SLC15A4 (MBL, Cat.#: BMP055, Lot #: 001, 1:250), rabbit anti-HA (Santa Cruz, Cat.#: sc-805, Lot #: H1215, 1:250).

Antibodies used in immunofluorescence or flow cytometry

Rabbit anti-HA (Cell Signaling, Cat.#: 3724, Lot #: 9, Dilution: 1:400), mouse anti-Lamp1 (Abcam, Cat.#: ab25630, Lot #: 792614, Dilution 1:200), Alexa Fluor 488-coupled goat anti-mouse (Thermo Fisher Scientific, Cat.#: A11001, Lot#: 2015565, Dilution: 1:400), Alexa Fluor 594-coupled goat anti-rabbit (Thermo Fisher Scientific, Cat.#: A11012, Lot#: 1892265, Dilution: 1:400), mouse anti-PD-L1-APC (Thermo Fisher Scientific, Cat.#: 17-5983-42, Lot #:4347834, Dilution: 1:20).

Validation

Specificity of custom SLC15A4-specific antibody (Genscript) and CXorf21/TASL antibodies (HPA001185, Sigma) was validated by western blot in SLC15A4/TASL knockout and overexpressing cells (see for example Fig. 2a, Fig. 4i, Extended Data Figure 3g). All other antibodies were bought from commercial vendors and validation for indicated species and applications can be found on the manufacturers website or the provided scientific citations on the same website.

Antibodies used in western blot and immunofluorescence:

rabbit anti-HA (Cell Signaling, Cat.#: 3724): <https://www.cellsignal.com/products/primary-antibodies/ha-tag-c29f4-rabbit-mab/3724>

mouse anti-Lamp1 (Abcam, Cat.#: ab25630): <https://www.abcam.com/lamp1-antibody-h4a3-ab25630.html>

Antibodies used in western blot:

rabbit anti-V5 (Cell Signaling, Cat.#: 13202): <https://www.cellsignal.com/products/primary-antibodies/v5-tag-d3h8q-rabbit-mab/13202>

rabbit anti-GFP (Cell Signaling, Cat.#: 2956): <https://www.cellsignal.com/products/primary-antibodies/gfp-d5-1-xp-rabbit-mab/2956>

rabbit anti-RAGA (Cell Signaling, Cat.#: 4357): <https://www.cellsignal.com/products/primary-antibodies/raga-d8b5-rabbit-mab/4357>

rabbit anti-TLR9 (Cell Signaling, Cat.#: 5845): <https://www.cellsignal.com/products/primary-antibodies/toll-like-receptor-9-d2c9-rabbit-mab/5845>

rabbit anti-TAK1 (Cell Signaling, Cat.#: 4505): <https://www.cellsignal.com/products/primary-antibodies/tak1-antibody/4505>

rabbit anti-IRAK4 (Cell Signaling, Cat.#: 4363): <https://www.cellsignal.com/products/primary-antibodies/irak4-antibody/4363>

rabbit anti-IKK α (Cell Signaling, Cat.#: 2682): <https://www.cellsignal.com/products/primary-antibodies/ikka-antibody/2682>

rabbit anti-IKK β (Cell Signaling, Cat.#: 8943): <https://www.cellsignal.com/products/primary-antibodies/ikkb-d30c6-rabbit-mab/8943>

rabbit anti-TBK1 (Cell Signaling, Cat.#: 51872): <https://www.cellsignal.com/products/primary-antibodies/tbk1-nak-e9h5s-mouse-mab/51872>

rabbit anti-IKK ϵ (Cell Signaling, Cat.#: 2905): <https://www.cellsignal.com/products/primary-antibodies/ikke-d20g4-rabbit-mab/2905>

rabbit anti-IRF3 (Cell Signaling, Cat.#: 11904): <https://www.cellsignal.com/products/primary-antibodies/irf-3-d6i4c-xp-rabbit-mab/11904>

rabbit anti-IRF7 (Cell Signaling, Cat.#: 4920): <https://www.cellsignal.com/products/primary-antibodies/irf-7-antibody/4920>

rabbit anti-phospho-IKK α / β Ser176/180 (Cell Signaling, Cat.#: 2697): <https://www.cellsignal.com/products/primary-antibodies/phospho-ikka-b-ser176-180-16a6-rabbit-mab/2697>

rabbit anti-phospho-IkBa Ser32/36 (Cell Signaling, Cat.#: 9246): <https://www.cellsignal.com/products/primary-antibodies/phospho-ikba-ser32-36-5a5-mouse-mab/9246>

rabbit anti-phospho-NF- κ B p65 Ser536 (Cell Signaling, Cat.#: 3033): <https://www.cellsignal.com/products/primary-antibodies/phospho-nf-kb-p65-ser536-93h1-rabbit-mab/3033>

rabbit anti-phospho-p38 MAPK Thr180/Tyr182 (Cell Signaling, Cat.#: 9211): <https://www.cellsignal.com/products/primary-antibodies/phospho-p38-mapk-thr180-tyr182-antibody/9211>

rabbit anti-phospho-SAPK/JNK Thr183/Tyr185 (Cell Signaling, Cat.#: 4668): <https://www.cellsignal.com/products/primary-antibodies/phospho-sapk-jnk-thr183-tyr185-81e11-rabbit-mab/4668>

rabbit phospho-STAT1 Y701 (Cell Signaling, Cat.#: 7649): <https://www.cellsignal.com/products/primary-antibodies/phospho-stat1-tyr701-d4a7-rabbit-mab/7649>

rabbit anti-CXorf21/TASL (Sigma, Cat.#: HPA001185): <https://www.sigmaaldrich.com/catalog/product/sigma/hpa001185>

rabbit anti-myc (Sigma, Cat.#: C3956): <https://www.sigmaaldrich.com/catalog/product/sigma/c3956>

mouse anti-GAPDH (Santa Cruz, Cat.#: sc-365062): <https://www.scbt.com/p/gapdh-antibody-g-9>

rabbit anti-IkB- α (Santa Cruz, Cat.#: sc-371): <https://www.scbt.com/p/ikappab-alpha-antibody-c-21>

mouse anti-LAMP2 (Santa Cruz, Cat.#: sc-18822): <https://www.scbt.com/p/lamp-2-antibody-h4b4>

mouse anti-STAT1 α (Santa Cruz, Cat.#: sc-417): <https://www.scbt.com/p/stat1alpha-p91-antibody-c-111>

rabbit anti-Actin (Cytoskeleton, Cat.#: AAN01-A): <https://www.cytoskeleton.com/aan01/>

mouse anti-Tubulin (Abcam, Cat.#: ab7291): <https://www.abcam.com/alpha-tubulin-antibody-dm1a-loading-control-ab7291.html>

rabbit anti-IRF5 (Abcam, Cat.#: ab181553): <https://www.abcam.com/irf5-antibody-epr17067-ab181553.html>

rabbit anti-TLR8 (MBL, Cat.#: pd047): Validated in Fig. 6d using TLR8 knockout cells and in Ishii N et al., J Immunol., 2014, PMID: 25297876

Antibodies used in immunoprecipitation

rabbit anti-SLC38A9 (Sigma, Cat.#: HPA043785): <https://www.sigmaaldrich.com/catalog/product/sigma/hpa043785> and validated in Ref. 24, (Rebsamen M et al., Nature, 2015)

rabbit anti-SLC15A4 (MBL, Cat.#: BMP055): Validated in Fig. 1h using SLC15A4 knockout cells

rabbit anti-HA (Santa Cruz, Cat.#: sc-805): <https://www.scbt.com/p/ha-probe-antibody-y-11>

Antibodies used in immunofluorescence.

Alexa Fluor 488-coupled goat anti-mouse IgG (H+L) (Life Technologies, Cat.#: A11001): <https://www.thermofisher.com/antibody/product/Goat-anti-Mouse-IgG-H-L-Cross-Adsorbed-Secondary-Antibody-Polyclonal/A-11001>

Alexa Fluor 594-coupled goat anti-rabbit IgG (H+L) (Life Technologies, Cat.#: A11012): <https://www.thermofisher.com/antibody/product/Goat-anti-Rabbit-IgG-H-L-Cross-Adsorbed-Secondary-Antibody-Polyclonal/A-11012>

Antibody used in Flow cytometry.

anti-PD-L1-APC (Thermo Fisher Scientific: 17-5983-42): <https://www.thermofisher.com/antibody/product/CD274-PD-L1-B7-H1-Antibody-clone-MIH1-Monoclonal/17-5983-42>

Eukaryotic cell lines

Policy information about [cell lines](#)

Cell line source(s)	HEK293T cells and THP1 cells were purchased from ATCC, THP1 DUAL™ reporter cell lines from Invivogen. CAL-1 cells were kindly provided by Prof. Takahiro Maeda, KBM-7 cells by Prof. Thijn Brummelkamp.
Authentication	All cell lines used in this study were authenticated by STR profiling, except CAL-1 as they were obtained directly from Prof. Maeda (Ref. 40).
Mycoplasma contamination	Cell lines used were tested negatively for mycoplasma contamination.
Commonly misidentified lines (See ICLAC register)	None of commonly misidentified cell lines were used in this study.

Flow Cytometry

Plots

Confirm that:

- ☒ The axis labels state the marker and fluorochrome used (e.g. CD4-FITC).
- ☒ The axis scales are clearly visible. Include numbers along axes only for bottom left plot of group (a 'group' is an analysis of identical markers).
- ☐ All plots are contour plots with outliers or pseudocolor plots.
- ☐ A numerical value for number of cells or percentage (with statistics) is provided.

Methodology

Sample preparation	For PD-L1, cells were washed with PBS and stained for 20 min. with APC-conjugated anti-PD-L1 antibodies (17-5983-42, Thermo Fisher Scientific, 1:20). For uptake assays of FITC-labeled CpG, cells were incubated with 1 μ M CpG-A (ODN2216) or CpG-B (ODN2006) for 0-120 minutes. Cells were washed with PBS and analyzed immediately by flow cytometry. For detection TASL-GFP-expressing cells, cells were washed with PBS and analyzed immediately by flow cytometry. For phagocytosis assays, cells were washed twice with ice cold PBS, harvested by scraping and resuspended in FACS buffer (2% FCS in ice cold PBS, 2 mM EDTA).
Instrument	BD Biosciences FACSCalibur; BD LSR Fortessa II
Software	Acquisition: BD CellQuest Pro v6.0; BD FACS DIVA v8.0.1; analysis: FlowJo v10
Cell population abundance	<i>Describe the abundance of the relevant cell populations within post-sort fractions, providing details on the purity of the samples and how it was determined.</i>
Gating strategy	Live cells were gated based on FSC/SSC.

- ☐ Tick this box to confirm that a figure exemplifying the gating strategy is provided in the Supplementary Information.

Structure and catalytic mechanism of a human triacylglycerol-synthesis enzyme

<https://doi.org/10.1038/s41586-020-2289-6>

Received: 11 December 2019

Accepted: 17 March 2020

Published online: 13 May 2020



Xuewu Sui^{1,2}, Kun Wang^{1,2}, Nina L. Gluchowski^{1,2,3}, Shane D. Elliott^{1,2}, Maofu Liao^{2,6}✉, Tobias C. Walther^{1,2,4,5,6}✉ & Robert V. Farese Jr^{1,2,4,6}✉

Triacylglycerols store metabolic energy in organisms and have industrial uses as foods and fuels. Excessive accumulation of triacylglycerols in humans causes obesity and is associated with metabolic diseases¹. Triacylglycerol synthesis is catalysed by acyl-CoA diacylglycerol acyltransferase (DGAT) enzymes^{2–4}, the structures and catalytic mechanisms of which remain unknown. Here we determined the structure of dimeric human DGAT1, a member of the membrane-bound *O*-acyltransferase (MBOAT) family, by cryo-electron microscopy at approximately 3.0 Å resolution. DGAT1 forms a homodimer through N-terminal segments and a hydrophobic interface, with putative active sites within the membrane region. A structure obtained with oleoyl-CoA substrate resolved at approximately 3.2 Å shows that the CoA moiety binds DGAT1 on the cytosolic side and the acyl group lies deep within a hydrophobic channel, positioning the acyl-CoA thioester bond near an invariant catalytic histidine residue. The reaction centre is located inside a large cavity, which opens laterally to the membrane bilayer, providing lipid access to the active site. A lipid-like density—possibly representing an acyl-acceptor molecule—is located within the reaction centre, orthogonal to acyl-CoA. Insights provided by the DGAT1 structures, together with mutagenesis and functional studies, provide the basis for a model of the catalysis of triacylglycerol synthesis by DGAT.

The triacylglycerol biosynthesis pathway was elucidated nearly 60 years ago⁵, and the DGAT enzymes were identified more than 20 years ago^{2–4}. Mammalian triacylglycerol synthesis is catalysed by DGAT1 (Extended Data Fig. 1a) and DGAT2 enzymes, which belong to different protein families with distinct predicted membrane topologies⁶. Both enzymes utilize fatty acyl-CoA and diacylglycerol (DAG) substrates, although DGAT1 has a broader acyl-acceptor substrate specificity⁷ (Extended Data Fig. 1a). DGAT1 and DGAT2 have distinct roles in cells⁸ and physiology^{9,10}, but together account for most mammalian triacylglycerol synthesis¹¹. Despite these insights, the enzymatic mechanism of triacylglycerol synthesis has remained unknown.

We investigated DGAT1, a member of the MBOAT family, which includes enzymes that acylate proteins (for example, ghrelin, Wnt and hedgehog) or lipids (for example, DAG and sterols)¹². DGAT1 is a polytopic endoplasmic reticulum membrane protein^{2,13} with a conserved histidine residue (His415 in human DGAT1) that is probably involved in catalysis^{13–16}. A crystal structure of DltB, a prokaryotic MBOAT, was reported recently¹⁷. However, DltB catalyses the transfer of an alanyl moiety to teichoic acid, and therefore its structure provides limited insights into the mechanism of lipid-acylating MBOATs.

Structure determination of human DGAT1

We determined the structure of DGAT1 by cryo-electron microscopy (cryo-EM). We purified human DGAT1 in digitonin, and

gel-filtration chromatography revealed two DGAT1-containing peaks, presumably representing different oligomeric states (Extended Data Fig. 1b, c). Initial studies of the major peak (peak 2 in Extended Data Fig. 1b) failed to yield high-resolution structures. We therefore reconstituted DGAT1 from the two peak fractions (Extended Data Fig. 1b, c) into PMAL-C8 amphipol (Extended Data Fig. 1d–f). Gel-filtration analysis revealed that the reconstituted DGAT1 also displayed two peaks (red and blue arrows, Extended Data Fig. 1d). The later elution peak was found to contain DGAT1 dimers; the earlier peak probably represented DGAT1 tetramers (Extended Data Fig. 1g), which have been observed previously^{13,14}. These two DGAT1 populations showed comparable activities (Extended Data Fig. 1f).

Reconstitution with amphipol improved homogeneity of DGAT1 particles (Extended Data Fig. 1g), enabling us to use DGAT1 in the later elution peak (blue arrow, Extended Data Fig. 1d) to generate a cryo-EM three-dimensional (3D) reconstruction at approximately 3.0 Å overall resolution (Extended Data Figs. 2, 3, Extended Data Table 1). This cryo-EM map showed well-defined side-chain densities for most amino acid residues, enabling unambiguous model building (Extended Data Fig. 3d). The first 65 amino acids and a loop region (amino acids 229–238) were invisible, probably owing to flexibility in these regions. A belt-shaped density of PMAL-C8 surrounded most of DGAT1 (Fig. 1a, b, Extended Data Fig. 2d), consistent with DGAT1 being mostly embedded in the membrane¹³.

¹Department of Molecular Metabolism, Harvard T. H. Chan School of Public Health, Boston, MA, USA. ²Department of Cell Biology, Harvard Medical School, Boston, MA, USA. ³Department of Gastroenterology, Hepatology, and Nutrition, Boston Children's Hospital, Boston, MA, USA. ⁴Broad Institute of MIT and Harvard, Cambridge, MA, USA. ⁵Howard Hughes Medical Institute, Boston, MA, USA. ⁶These authors contributed equally: Maofu Liao, Tobias C. Walther, Robert V. Farese Jr. ✉e-mail: maofu_liao@hms.harvard.edu; twalther@hsph.harvard.edu; robert@hsph.harvard.edu

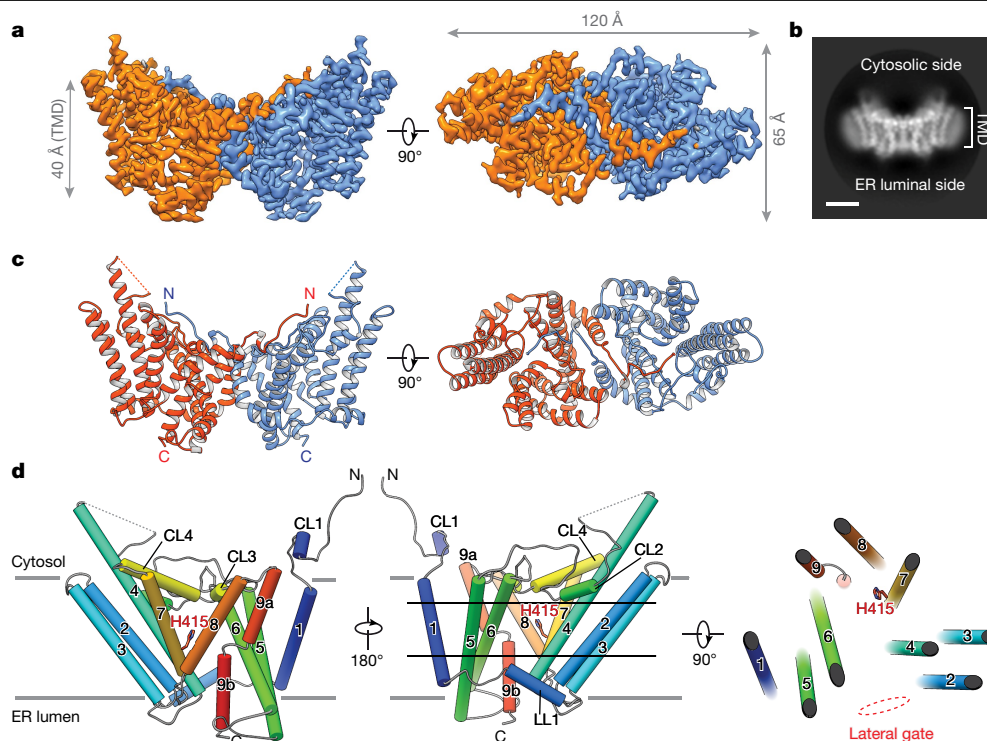


Fig. 1 | Cryo-EM structure of human DGAT1. **a**, Cryo-EM map of human DGAT1, coloured by monomer, viewed along the membrane plane or from the cytosolic side. Map is contoured at 6σ . TMD, transmembrane domain. **b**, Representative two-dimensional (2D) average of DGAT1 cryo-EM images. A micelle-like density surrounds the major portion of DGAT1. Scale bar, 40 Å. ER, endoplasmic

reticulum. **c**, Ribbon representation of the human DGAT1 dimer. The dashed line indicates a disordered segment (residues 229–238) not resolved in the cryo-EM map. **d**, Topology of DGAT1 monomer. The conserved His415 residue is shown. The red dashed oval (right) indicates the membrane-embedded lateral gate to the central cavity.

Overall architecture of human DGAT1

The cryo-EM map reveals that DGAT1 forms a dimer (Fig. 1a–c). Each DGAT1 subunit has nine transmembrane helices, with N and C termini located on the cytosolic and luminal sides of the endoplasmic reticulum membrane, respectively (Fig. 1c, d). Short helices in cytosolic loop (CL) and luminal loop (LL) regions, including CL2–CL4 and LL1, orient in parallel to the membrane surface (Fig. 1d). The DGAT1 dimer forms through extensive hydrogen-bonding interactions of the first 20 resolved residues (His69–Gly87) with a cytosolic surface groove on the opposing subunit, and through hydrophobic interactions of the transmembrane helix 1 (TM1) region of Phe82–Ile98 with the other monomer (Extended Data Fig. 4a). Additional densities, the shapes and sizes of which are consistent with four phospholipids, were present at the dimer interface and appear to contribute to the contacts between the DGAT1 monomers (Extended Data Fig. 4b). Consistently, phospholipids were identified in extracts of purified DGAT1 after amphipol reconstitution (Extended Data Fig. 4c).

To assess the functional importance of dimerization, we measured acyl-transferase activities in lysates of N-terminal truncation mutants expressed in cells lacking DGAT1 (Extended Data Fig. 4d, e). Similar to previous reports^{13,15}, deletion of the first 85 ($\Delta 85$) or 90 residues of DGAT1 led to lower protein expression levels and reduced, but still detectable, activity (Extended Data Fig. 4f–h). The purified detergent-solubilized $\Delta 85$ -mutant protein exhibited almost complete absence of enzyme activity (Extended Data Fig. 10c), indicating that the N-terminal region is required for optimal activity.

The DGAT1 active site is within the membrane

The transmembrane helices of DGAT1 form a large central cavity within the membrane that is open to the bilayer via a wide lateral gate (Fig. 2a, b)

lined by TM2, TM4, TM5, TM6 and LL1 (Figs. 1d, 2b). The conserved catalytic histidine residue (His415)¹³ is buried inside the central cavity (Figs. 1d, 2d). As well as through the lateral gate, this cavity is accessible through openings on the cytosolic and luminal sides (Fig. 2a, c).

Structural comparison of DGAT1 and DltB¹⁷ reveals an overall similar fold, in which the majority of transmembrane helices form a concave-shaped ridge on either side of the membrane and the conserved histidine residue is located in the central region embedded in the membrane (Fig. 2b, c). A DALI search with the DGAT1 structure showed that this architecture appears to be unique to DGAT1 and DltB. We therefore refer to this conserved domain structure as the ‘MBOAT core’ (Extended Data Fig. 5a–c). Within the MBOAT core, a tunnel-like region equivalent to the alanyl-donor binding pocket of DltB is also present in DGAT1 (Fig. 2b, green circle). Despite their overall similarity, the lateral gate found in DGAT1 is absent from DltB. Instead, the N-terminal transmembrane helices TM1 and TM2 of DltB are in close proximity, blocking lateral access to the central cavity from the membrane (Fig. 2b, Extended Data Figs. 5a, b). This suggests that the acyl-acceptor lipid substrates for DGAT1, such as DAG, access the active site through the lateral gate within the membrane, whereas DltB, which alanylates extracellular teichoic acid, does not require a membrane opening for substrates. This hypothesis is consistent with a computationally predicted structure of ghrelin O-acyltransferase¹⁸, an MBOAT that acylates the secreted peptide hormone ghrelin at the luminal side of the endoplasmic reticulum, in which the transmembrane region lacks an opening within the lipid bilayer.

The putative active site of DGAT1 contains both the invariant His415 and several highly conserved nearby polar residues, including Asn378, Gln437 and Gln465, with their side chains oriented towards the cavity centre (Fig. 2d, Extended Data Fig. 6a). The orientation of His415 appears to be stabilized by a hydrogen bond of its N δ_1 atom with the S δ atom of Met434 (Fig. 2d). Single alanine mutations of His415 or Asn378 abolished DGAT1 activity, whereas Gln437Ala, Met434Ala or Gln465Ala mutations

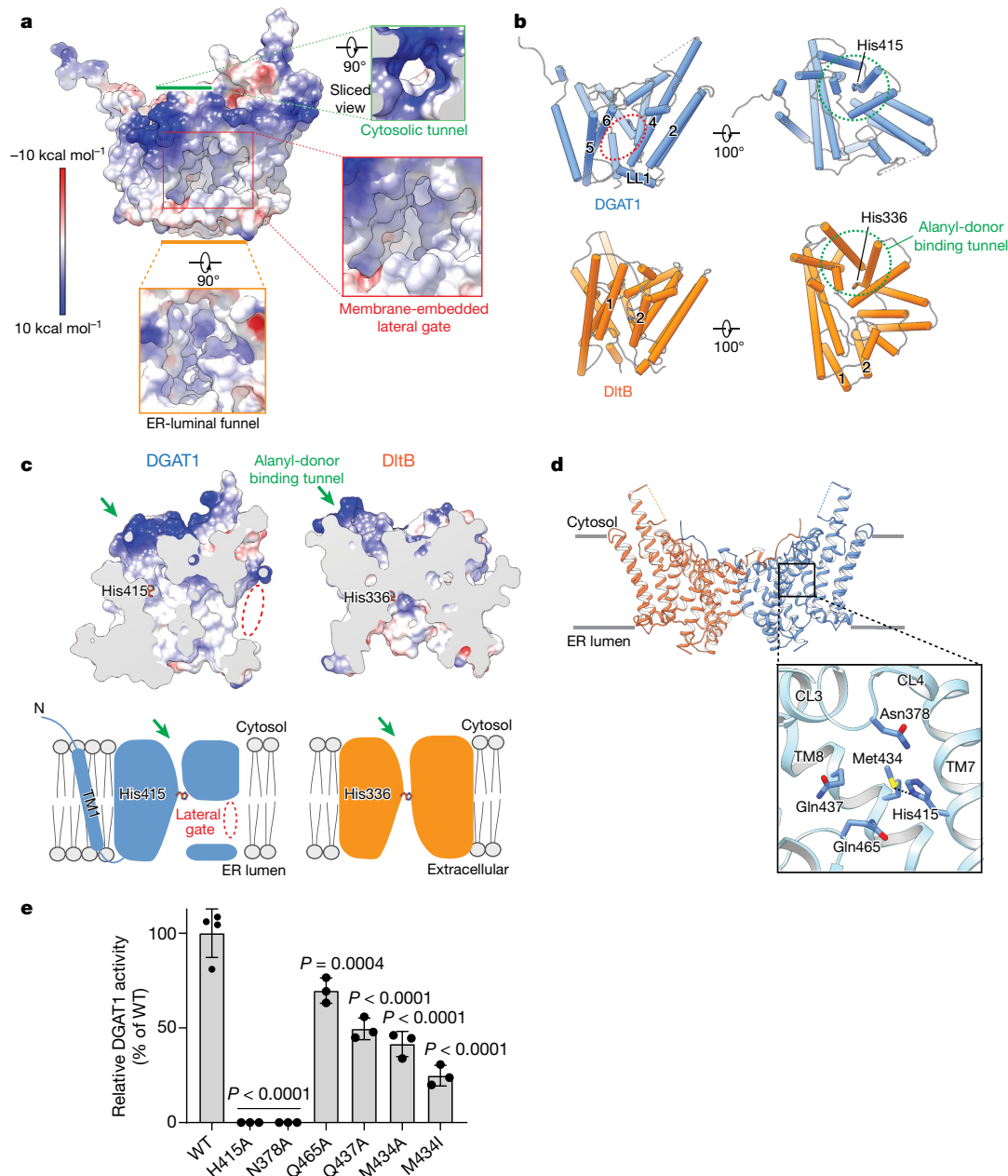


Fig. 2 | DGAT1 reaction centre and structural comparison with DltB. a, Three major openings leading to the DGAT1 putative catalytic centre include a cytosolic tunnel (green box), an ER-luminal funnel (orange box), and a membrane-embedded lateral gate (red box). The electrostatic surface is shown. **b,** DGAT1 structural comparison with DltB. DGAT1 and DltB structures are superimposed and separately shown as cylinders. The lateral gate in DGAT1 (red dashed oval) is absent in DltB owing to the presence of TM1 and TM2, which block membrane access to the catalytic centre. A cytosolic tunnel in DGAT1 (green box in **a**) is also present in DltB and functions as a binding tunnel for alanyl-donor protein DltC (green dashed circle). **c,** Sliced-surface views (top) of superimposed DGAT1 and DltB as in **b**, and cartoon representations (bottom) illustrating the common and distinct structural features. The red oval indicates

reduced activity by 50–75% (Fig. 2e). Together, these studies suggest that the central cavity of DGAT1 within the membrane is the catalytic site.

Acyl-CoA binding to DGAT1

To investigate DGAT1-mediated acyl-CoA binding and acyl-chain transfer, we characterized the structure of human DGAT1 in the presence of

the lateral gate shown in **a**, and the green arrow indicates the cytosolic tunnel shown in **a** and **b**. Note that the N terminus and TM1 of DGAT1 appear as extra structural elements compared with DltB (also see Extended Data Fig. 5a, b). **d,** Zoomed-in view of the DGAT1 active centre. The conserved His415 and highly conserved polar residues in the vicinity are shown as sticks. A hydrogen bond between His415 and Met434 (3.1 Å) is shown as a dashed line. **e,** DGAT1 activities of wild type (WT) and reaction centre mutants expressed in DGAT1-knockout cells. Activities were normalized to the amount of DGAT1 protein expressed. Data are mean \pm s.d.; $n = 4$ independent experiments for WT, $n = 3$ independent experiments for all mutants. One-way ANOVA with Dunnett's post hoc test was applied.

oleoyl-CoA, a preferred substrate³, and generated cryo-EM maps at 3.2 Å and 2.8 Å resolution (Extended Data Figs. 7, 8). These two maps are essentially identical for DGAT1 protein and show strong densities for the bound oleoyl-CoA (Fig. 3a, Extended Data Fig. 9a). In the 3.2 Å resolution map, a complete oleoyl-CoA molecule is visible (Fig. 3a), representing an unprocessed substrate. In the cryo-EM maps calculated with other particle images, the oleoyl-CoA density appears to be broken

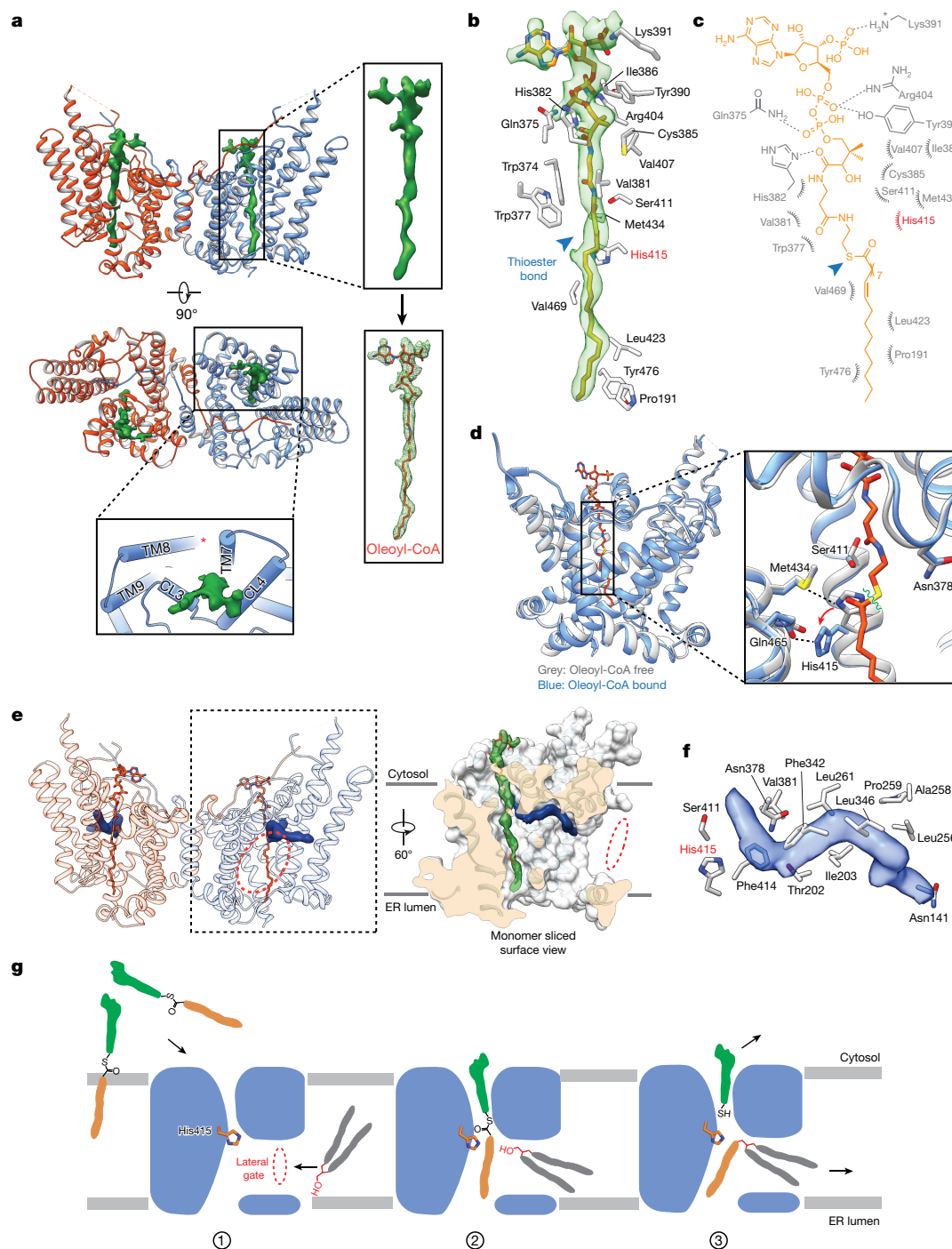


Fig. 3 | Cryo-EM structures of DGAT1 complexed with a fatty acyl-donor substrate and a lipid-like molecule resembling an acyl-acceptor substrate.

a, Cryo-EM structure of human DGAT1 bound to an intact oleoyl-CoA molecule (green density at 3.5 σ contour level). Zoomed-in view shows the modelled oleoyl-CoA molecule superimposed with the electron microscopy density. Lower panel illustrates the cytosolic tunnel of DGAT1, consisting of TM7, TM8, TM9, CL3 and CL4, which interacts with oleoyl-CoA substrate. The red asterisk indicates the gap between TM7 and TM8 in the cytosolic membrane leaflet.

b, c, Interactions between oleoyl-CoA and DGAT1. Residues interacting with oleoyl-CoA within a 4 Å distance are shown. Detailed molecular interactions are depicted in **c**. Polar interactions (dashed lines) and interaction distances are labelled. Non-dipolar interactions are shown as spiked arcs. The blue arrowhead indicates the thioester bond in the acyl-CoA substrate.

d, Conformational changes in the DGAT1 active centre upon oleoyl-CoA binding. The red arrow indicates the His415 conformational change before (grey) and after (blue) oleoyl-CoA binding, and dashed lines indicate the

hydrogen bonds between His415 and Met434 without oleoyl-CoA and between His415 and Gln465 (2.4 Å) upon binding to oleoyl-CoA. Asn378, which is essential for DGAT1 activity, is also shown. The green wavy line indicates the scissile bond in the thioester group during the acyl-transfer reaction. **e**, Left, unidentified lipid-shaped density (blue) residing in the DGAT1 central cavity. The lateral gate (red dashed circle) and oleoyl-CoA molecules (sticks) are shown. Right, surface representation, showing the orientation of the lipid-like density in the monomer of the DGAT1-acyl-CoA complex structure. **f**, Residues accommodating the lipid-like density (transparent blue surface) shown in the acyl-CoA free DGAT1 structure. **g**, Hypothetical model for DGAT1-catalysed triacylglycerol formation. The reaction is initiated by binding of acyl-donor (via the cytosolic channel) and acyl-acceptor (via the lateral gate) substrates to the active centre within the membrane (step 1). The acyl-transfer reaction is catalysed by the conserved His415 residue (step 2). The reaction product of free CoA-SH is released into the cytosol, and the hydrophobic triacylglycerol molecule diffuses into the lipid bilayer (step 3).

at the thioester bond (Extended Data Fig. 9a), possibly representing a state after catalytic cleavage. However, validation that the cleaved acyl-CoA is a post-reaction state requires further investigation.

Our structure of DGAT1 bound to intact oleoyl-CoA reveals that the acyl-CoA molecule occupies the cytosolic tunnel, with the CoA moiety at the cytosolic face and its acyl chain extending through the reaction centre towards the endoplasmic reticulum lumen (Fig. 3a). This orientation of the acyl-CoA is consistent with acyl-CoA molecules being synthesized on the cytosolic side of the endoplasmic reticulum membrane¹⁹. Acyl-CoA and DGAT1 interaction is mediated predominantly through the 4'-phosphopantetheine moiety in the oleoyl-CoA molecule (Fig. 3b, c). The adenosine group of CoA contributes minimally to interactions with DGAT1. The relatively weak cryo-EM density in this region suggests that the adenosine is mobile and could possibly adopt various conformations (Extended Data Fig. 9b). By contrast, the density for the 4'-phosphopantetheine moiety and surrounding residues is clearly visible (Extended Data Fig. 9b). Single alanine mutations of the residues interacting with oleoyl-CoA decreased DGAT1 activity levels, although by less than 50% (Extended Data Fig. 9c). More bulky hydrophobic side-chain mutations, intended to block the binding of acyl-CoA to the tunnel (Val381Trp, Cys385Trp, Val407Phe and Ser411Ile) decreased DGAT1 activity, and a Ser411Trp mutation, which we predicted would block the tunnel, completely inactivated DGAT1 (Extended Data Fig. 9d). The loss of activity was apparently not caused by protein misfolding (Extended Data Fig. 10a, b). Activity was also validated by using purified DGAT1 (Val381Trp and Ser411Trp) mutants (Extended Data Fig. 10c).

The oleoyl-CoA in the tunnel is surrounded by TM7, TM8, TM9, CL3 and CL4, leaving a gap between TM7 and TM8 at the level of the cytosolic membrane leaflet (Fig. 3a, asterisk). This gap might accept membrane-partitioned acyl-CoA substrates, although soluble acyl-CoA could also enter the channel from the cytoplasm. The distal end of the acyl chain of oleoyl-CoA interacts with DGAT1 deep within the hydrophobic channel, suggesting that the binding of longer acyl chains may help to accurately position the acyl-donor substrate for the reaction. Consistently, DGAT1 was more active with longer acyl-chain substrates (C16–C20) than shorter ones (C10–C12) (Extended Data Fig. 9e). The enzyme exhibited activity for a variety of saturated and unsaturated long-chain fatty acyl-CoA molecules (Extended Data Fig. 9e), as reported^{3,20}. An analysis of the conservation of DGAT1 and selected MBOAT sequences showed that the acyl-CoA binding tunnel within the MBOAT core of DGAT1 is the most conserved region among MBOAT enzymes (Extended Data Fig. 6b, c). This tunnel is analogous to the pocket in DltB for binding DltC, the alanyl-donor protein¹⁷ (Fig. 2b, c, Extended Data Fig. 5b, c).

Comparison of the structures of DGAT1 with or without bound acyl-CoA reveals no large conformational differences (Fig. 3d), similar to the minor structural changes for DltB upon DltC binding¹⁷. However, several residues near the cleavage site of acyl-CoA, including Ser411, show altered conformations upon acyl-CoA binding. Among these, His415 flips towards the endoplasmic reticulum-luminal side when acyl-CoA binds (Fig. 3d). This change involves the breaking of the hydrogen bond between His415 and Met434, which opens the tunnel to accommodate the acyl-CoA substrate and enables the formation of a hydrogen bond between the His415 Ne₂ atom and the Gln465 Oe₁ atom (Fig. 3d), positioning His415 near the thioester bond of the acyl-CoA. Thus, acyl-CoA binding to DGAT1 results in small but important conformational changes in the active-site region that may prime the enzyme for catalysis.

A putative acyl acceptor

The cleavage site for oleoyl-CoA is presumably within a short distance of a lipid acceptor, such as DAG. Of note, our cryo-EM maps (with or without oleoyl-CoA) revealed a strong, elongated, lipid-like density in the central cavity (Fig. 3e), orthogonal to the bound oleoyl-CoA and

in close proximity to His415 (Fig. 3e, f). The shape of this density is consistent with a DAG that was co-purified with the enzyme (Extended Data Fig. 4c). Hydrophobic residues, including Phe342, Leu261, Val381 and the highly conserved Asn378, line the region and form a channel surrounding the lipid-like density (Fig. 3f). The channel exhibits a bent, hydrophobic pathway that appears to allow the binding of hydrophobic linear or curvilinear molecules (Extended Data Fig. 9g). The bent architecture of this tunnel may explain how DGAT1 distinguishes acyl acceptors such as DAG or long-chain alcohols⁷ from more planar and rigid substrates such as cholesterol, which is acylated by acyl CoA:cholesterol acyltransferases²¹. In the cryo-EM map of DGAT1 containing a cleaved oleoyl-CoA, the cleaved acyl chain is merged onto the orthogonally orientated density, giving rise to a density with a three-way branched shape resembling triacylglycerol (Extended Data Fig. 9f). Further studies are required to determine the identity of the lipid-like density and to elucidate the basis of acyl-acceptor preferences of DGAT1.

Discussion

Our cryo-EM structures of human DGAT1, obtained in the absence and presence of bound fatty acyl-CoA substrate, reveal the detailed molecular architecture of this mammalian MBOAT and, together with extensive functional studies, provide mechanistic insight into the catalysis of lipid acylation mediated by DGAT1. Our results reveal the dimeric architecture of DGAT1, the arrangement of crucial residues within the catalytic centre, the structural basis for acyl-CoA binding and thioester cleavage, and a potential mechanism of acyl transfer to a lipid acceptor.

On the basis of our results, we propose a model for triacylglycerol synthesis mediated by DGAT1 in which an acyl-CoA substrate binds, from the cytosolic side, via a combination of electrostatic features of the cytosolic face and a deep, hydrophobic tunnel (Fig. 3b, c). When an acyl acceptor—such as DAG—binds in an orthogonal hydrophobic tunnel accessed from within the lipid bilayer, the acyl chain from acyl-CoA is transferred to the acceptor in a reaction involving the conserved His415 residue²², ultimately generating CoA-SH and triacylglycerol (see model, Fig. 3g). The crucial Asn378 residue may serve to interact with the acyl-acceptor substrate. The reaction products are released to either the cytoplasm (CoA-SH) or the membrane bilayer (triacylglycerol); within the membrane bilayer, the triacylglycerol can initiate lipid-droplet formation²³. Additional structural and biochemical studies are needed to test this model and delineate the catalytic mechanism.

Our structure shows that the N-terminal region of DGAT1 mediates dimerization, whereas the analogous region of the DltB monomer instead blocks the lateral gate to the membrane. This suggests evolutionary functional adaptations that occurred within MBOATs to accommodate different substrates. Previous studies indicated that the N-terminal region of DGAT1 mediates its oligomerization^{13,15} and may be involved in acyl-CoA binding and enzyme regulation^{15,16}. Because our structure lacks N-terminal segments, it does not provide new insights into these possibilities.

Our DGAT1 structure provides a framework for understanding mutations in human *DGAT1*, which cause congenital diarrhoea²⁴. Most of these *DGAT1* mutations result in a loss of protein expression, but as additional mutations are identified, they can be mapped onto the enzyme structure to predict the functional consequences. The DGAT1 structure may also further our understanding of the mechanisms of DGAT1 inhibitors that have been developed for treating metabolic diseases^{25,26}.

Online content

Any methods, additional references, Nature Research reporting summaries, source data, extended data, supplementary information, acknowledgements, peer review information; details of author contributions and competing interests; and statements of data and code availability are available at <https://doi.org/10.1038/s41586-020-2289-6>.

1. Cohen, J. C., Horton, J. D. & Hobbs, H. H. Human fatty liver disease: old questions and new insights. *Science* **332**, 1519–1523 (2011).
2. Cases, S. et al. Identification of a gene encoding an acyl CoA:diacylglycerol acyltransferase, a key enzyme in triacylglycerol synthesis. *Proc. Natl Acad. Sci. USA* **95**, 13018–13023 (1998).
3. Cases, S. et al. Cloning of DGAT2, a second mammalian diacylglycerol acyltransferase, and related family members. *J. Biol. Chem.* **276**, 38870–38876 (2001).
4. Lardizabal, K. D. et al. DGAT2 is a new diacylglycerol acyltransferase gene family: purification, cloning, and expression in insect cells of two polypeptides from *Mortierella ramanniana* with diacylglycerol acyltransferase activity. *J. Biol. Chem.* **276**, 38862–38869 (2001).
5. Weiss, S. B., Kennedy, E. P. & Kiyasu, J. Y. The enzymatic synthesis of triglycerides. *J. Biol. Chem.* **235**, 40–44 (1960).
6. Yen, C. L., Stone, S. J., Koliwad, S., Harris, C. & Farese, R. V. Jr. Thematic review series: glycerolipids. DGAT enzymes and triacylglycerol biosynthesis. *J. Lipid Res.* **49**, 2283–2301 (2008).
7. Yen, C. L., Monetti, M., Burri, B. J. & Farese, R. V. Jr. The triacylglycerol synthesis enzyme DGAT1 also catalyzes the synthesis of diacylglycerols, waxes, and retinyl esters. *J. Lipid Res.* **46**, 1502–1511 (2005).
8. Wilfling, F. et al. Triacylglycerol synthesis enzymes mediate lipid droplet growth by relocalizing from the ER to lipid droplets. *Dev. Cell* **24**, 384–399 (2013).
9. Chitraju, C. et al. Triglyceride synthesis by DGAT1 protects adipocytes from lipid-induced ER stress during lipolysis. *Cell Metab.* **26**, 407–418 (2017).
10. Chitraju, C., Walther, T. C. & Farese, R. V. Jr. The triglyceride synthesis enzymes DGAT1 and DGAT2 have distinct and overlapping functions in adipocytes. *J. Lipid Res.* **60**, 1112–1120 (2019).
11. Stone, S. J. et al. Lipopenia and skin barrier abnormalities in DGAT2-deficient mice. *J. Biol. Chem.* **279**, 11767–11776 (2004).
12. Hofmann, K. A superfamily of membrane-bound O-acyltransferases with implications for Wnt signaling. *Trends Biochem. Sci.* **25**, 111–112 (2000).
13. McFie, P. J., Stone, S. L., Banman, S. L. & Stone, S. J. Topological orientation of acyl-CoA:diacylglycerol acyltransferase-1 (DGAT1) and identification of a putative active site histidine and the role of the N terminus in dimer/tetramer formation. *J. Biol. Chem.* **285**, 37377–37387 (2010).
14. Cheng, D. et al. Human acyl-CoA:diacylglycerol acyltransferase is a tetrameric protein. *Biochem. J.* **359**, 707–714 (2001).
15. Caldo, K. M. P. et al. Diacylglycerol acyltransferase 1 is regulated by its N-terminal domain in response to allosteric effectors. *Plant Physiol.* **175**, 667–680 (2017).
16. Panigrahi, R. et al. Intrinsic disorder in the regulatory N-terminal domain of diacylglycerol acyltransferase 1 from *Brassica napus*. *Sci. Rep.* **8**, 16665 (2018).
17. Ma, D. et al. Crystal structure of a membrane-bound O-acyltransferase. *Nature* **562**, 286–290 (2018).
18. Campaña, M. B. et al. The ghrelin O-acyltransferase structure reveals a catalytic channel for transmembrane hormone acylation. *J. Biol. Chem.* **294**, 14166–14174 (2019).
19. Grevengoed, T. J., Klett, E. L. & Coleman, R. A. Acyl-CoA metabolism and partitioning. *Annu. Rev. Nutr.* **34**, 1–30 (2014).
20. Caldo, K. M., Greer, M. S., Chen, G., Lemieux, M. J. & Weselake, R. J. Purification and properties of recombinant *Brassica napus* diacylglycerol acyltransferase 1. *FEBS Lett.* **589**, 773–778 (2015).
21. Chang, T. Y., Li, B. L., Chang, C. C. & Urano, Y. Acyl-coenzyme A:cholesterol acyltransferases. *Am. J. Physiol. Endocrinol. Metab.* **297**, E1–E9 (2009).
22. Röttig, A. & Steinbüchel, A. Acyltransferases in bacteria. *Microbiol. Mol. Biol. Rev.* **77**, 277–321 (2013).
23. Walther, T. C., Chung, J. & Farese, R. V. Jr. Lipid droplet biogenesis. *Annu. Rev. Cell Dev. Biol.* **33**, 491–510 (2017).
24. Haas, J. T. et al. DGAT1 mutation is linked to a congenital diarrheal disorder. *J. Clin. Invest.* **122**, 4680–4684 (2012).
25. Cao, J. et al. Targeting Acyl-CoA:diacylglycerol acyltransferase 1 (DGAT1) with small molecule inhibitors for the treatment of metabolic diseases. *J. Biol. Chem.* **286**, 41838–41851 (2011).
26. He, S. et al. Discovery of a potent and selective DGAT1 inhibitor with a piperidinyloxy-cyclohexanecarboxylic acid moiety. *ACS Med. Chem. Lett.* **5**, 1082–1087 (2014).

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

© The Author(s), under exclusive licence to Springer Nature Limited 2020

Methods

No statistical methods were used to predetermine sample size. The experiments were not randomized. The investigators were not blinded to allocation during experiments and outcome assessment.

Protein expression and purification

DGAT1 from *Homo sapiens* (UniPort ID: O75907) was expressed in HEK293 GnTi- (ATCC) cells using the BacMam system. In brief, the DGAT1 cDNA sequence was cloned into a pFastBacMam vector constructed with the mammalian CMV promoter. The protein was fused with an N-terminal His₈ tag followed by a maltose-binding protein (MBP). A TEV cleavage site was inserted between the MBP and DGAT1 sequences. The bacmid for expressing the MBP-fusion protein was generated in DH10Bac *Escherichia coli* cells (Thermo Fisher Scientific) by transformation of the construct. Baculovirus was generated by transfecting *Spodoptera frugiperda* (Sf9) cells (Expression Systems) cultured at 27 °C in ESF921 medium (Expression Systems) with the bacmid using the BaculoPORTER transfection reagent (Genlantis). After that, the baculovirus was amplified twice in Sf9 cells to obtain sufficient virus for large-scale infection. HEK293 GnTi- cells were grown in suspension at 37 °C in FreeStyle 293 Expression Medium (Thermo Fisher Scientific) supplemented with 1% FBS and 1 × Glutamax solution (Gibco). When the cells reached a density of $2.0\text{--}2.5 \times 10^6$ cells per ml, baculovirus was added to the culture at 5% (v/v). After 15 h infection at 37 °C, the culture was supplemented with 10 mM sodium butyrate to boost the expression. After further growth for ~36 h, cells were collected and washed once in PBS, and cell pellets were flash frozen in liquid nitrogen and stored at –80 °C or placed on ice for immediate use.

All protein purification steps were performed at 4 °C. Cell pellets collected from 1 l cell culture were resuspended in 80 mL TSGE buffer (50 mM Tris-HCl, pH 8.0, 400 mM NaCl, 10% (v/v) glycerol and 1 mM EDTA) and supplemented with 1 × complete protease inhibitor cocktail (Roche). To lyse the cells, GDN detergent (Anatrace) was added into the solution at a final concentration of 0.5% (w/v), and the mixture was generally agitated for 1 h before MgCl₂ was added to a final concentration of 5 mM. The mixture was agitated for 1 more hour. Insoluble debris was removed by centrifugation, and the supernatant was incubated with prewashed amylose resin for 2 h. The resin was carefully washed to remove contaminant proteins. The MBP-fused DGAT1 protein was eluted in TSGE buffer containing 20 mM maltose and 0.1% (w/v) digitonin (Sigma-Aldrich). The eluate containing MBP-DGAT1 was concentrated, and the MBP tag was cleaved by TEV protease. The protein was then further purified by size-exclusion chromatography on a Superose 6 column (GE Healthcare) in TSM buffer (50 mM Tris-HCl, pH 7.5, 400 mM NaCl, 10 mM MgCl₂) supplemented with 0.05% (w/v) digitonin. The peak fractions containing DGAT1 were collected and concentrated. For preparation of samples for cryo-EM analysis, the purified DGAT1 in detergent was mixed with PMAL-C8 (Anatrace) at a 1:3–1:5 (w/w) ratio, followed by gentle agitation overnight in cold room. Detergent was then removed with Bio-Beads SM-2 (Bio-Rad) for 1 h, and the beads were subsequently removed over a disposable polyprep column. The eluent was cleared by passage through a 0.22-μm filter before further purification by gel filtration with TSM buffer. The peak containing DGAT1 was collected for further use. For some assays of partially purified DGAT1 mutant proteins, recombinant proteins were transiently expressed in HEK293 cells and purified by the His₈ tag before activity measurements were performed.

Electron microscopy sample preparation and data acquisition

Negatively stained specimens were prepared using an established protocol with minor modifications²⁷. Specifically, 2.5 μl of purified DGAT1 in PMAL-C8 at 0.03–0.05 mg ml^{–1} were applied to glow-discharged copper electron microscopy grids covered with a thin layer of continuous carbon film, and the grids were stained with 1.5% (w/v) uranyl formate for 30 s. The grids were imaged on a Tecnai T12 microscope (Thermo

Fisher Scientific) operated at 120 kV and equipped with a 4k × 4k charge-coupled device camera (UltraScan 4000; Gatan). A nominal magnification of 67,000×, corresponding to a pixel size of 1.68 Å on the specimen, and a defocus of –1.5 μm were used to record the images.

For cryo-EM samples, 2–3 μl of purified DGAT1 in PMAL-C8 was applied to a Quantifoil holey carbon grid (Cu R1.2/1.3; 400 mesh) that was glow-discharged for 30 s. Optimal particle distribution was obtained with a protein concentration of 4–5 mg ml^{–1}. After applying protein, the grids were blotted with a Whatman #1 filter paper for 5 s with ~95% humidity at 4 °C and plunge frozen in liquid ethane cooled by liquid nitrogen using a Vitrobot Mark IV system (Thermo Fisher Scientific). For the oleoyl-CoA treated sample, oleoyl-CoA freshly prepared in water was added into the concentrated DGAT1 at 5.5 mg ml^{–1} in PMAL-C8 with a final concentration of 1 mM. The mixture was incubated on ice for 1 h before freezing the cryo-EM grids. Cryo-EM data were collected on a Titan Krios electron microscope (Thermo Fisher Scientific). Images were recorded using SerialEM²⁸. Refer to Supplementary Table 1 for the detailed data collection parameters.

Electron microscopy data processing

For negative-stain electron microscopy data, the images were binned over 2 × 2 pixels, yielding a pixel size of 3.36 Å, for further processing using Simplified Application Managing Utilities for EM labs (SAMUEL) scripts²⁹. For cryo-EM data, drift correction was performed using MotionCor2³⁰, and images were binned 2 × 2 by Fourier cropping to a pixel size of 3.3 Å. The defocus values were determined using CTFFIND4³¹ and motion-corrected sums without dose-weighting. Motion-corrected sums with dose-weighting were used for all other steps of image processing. Particle picking was performed using a semi-automated procedure³². The 2D classification of selected particle images were performed using samclasscas.py, which uses SPIDER operations to run 10 cycles of correspondence analysis, *k*-means classification, and multireference alignment, or RELION 2D classification^{33,34}. Initial 3D models were generated with 2D averages using SPIDER 3D projection matching refinement (samrefine.py) starting from a cylindrical density that mimics the general shape and size of DGAT1. The 3D classification and refinement were performed using relion3_refine_mpi in RELION. One round of 3D classification without applying symmetry was performed to remove bad particles. Subsequently, particle images that generated 2D class averages showing clear structural features were combined for one round of 2D classification followed by 3D refinement with C2 symmetry. Refer to Extended Data Figs. 2, 7 for the detailed data processing procedure. All refinements followed the gold-standard procedure in which two half datasets are refined independently. The overall resolutions were estimated based on the gold-standard Fourier shell correlation (FSC) = 0.143 criterion. Local resolution variation of cryo-EM maps was calculated using ResMap³⁵. The amplitude information of the final maps was corrected by applying a negative B factor using relion_postprocessing with the –auto_bfac option. The details related to data processing are summarized in Extended Data Table 1.

Model building and refinement

DGAT1 density maps in MRC/CCP4 format were converted to the structure factors MTZ format with map_to_structure_factors. Then a rough initial model was obtained in by using map_to_model module on the auto-sharpened map. All these functional modules were implemented in the PHENIX suite³⁶. The modelling building was then performed manually in COOT³⁷. The DGAT1 map shows clear densities for most transmembrane helices that allowed a confident sequence registration by aromatic and other residues with bulky side chains to these α-helical regions in the density map. The linker domains that connect the transmembrane-helical regions were then built by connecting helical domains, with large and well-resolved side chains serving as markers. After building the DGAT1 monomer model, the coordinate file was docked into DGAT1 dimer map in UCSF Chimera³⁸ to generate

Article

the dimer model. Then the dimer model was manually adjusted and refined in PHENIX real-space refinement package³⁶. The refined model was visually inspected and adjusted in COOT³⁷, and the resulting map was further put back through the real-space refinement procedure to undergo further refinement. After several cycles of manual model adjustments and computational refinement, the ligands for oleoyl-CoA and POPE lipids were placed into the residual active-site electron density. Atomic coordinates and geometric restraints for the ligands were generated using the GRADE server (Global Phasing). Then the protein coordinates with ligands were refined. This iterative refinement process was repeated until the dimer model reached optimal geometric statistics as evaluated by MolProbity³⁹. The regions show weak signals were omitted in the final model. The protein–protein and protein–ligand interactions in the structures were analysed by LigPlot⁴⁰.

Generation of DGAT1-knockout cells

The DGAT1-knockout SUM159 cell line in wild-type background was generated by CRISPR–Cas9 gene editing⁴¹. The sequence 5′-CCACGGTAGTTGCTGAAGCCACT-3′ was used as a guide RNA (gRNA) to direct Cas9 into exon 2 of the *DGAT1* locus. Cells were selected with 1.5 µg ml⁻¹ puromycin for 48 h. Genomic DNA of clones was amplified by PCR (sense: 5′-GAGAGCTTTGCCCACTGTAGG-3′, antisense: 5′-CTGGGTGACAGAGCCTGTTC-3′). PCR products were sequenced and revealed a clone (designated 2.10) with two deletion alleles, a 1-bp deletion (5′-AGTGGCTTCAGCAACTCCGTGGCATCCTGAACTGG-3′ → 5′-AGTGGCTTCAGCAACTCCGTGGCATCCTGAACTGG-3′) and predicted to result in a 14-bp deletion (5′-AGTGGCTTCAGCAACTACCGTGGCATCCTGAACTGG-3′ → 5′-AGTGGCTTCAGCAACTGAACTGG-3′). These deletions (underlined bases) result in frameshifts and premature stop codons within the first 100 amino acids. The knockout was verified by immunoblotting.

DGAT1 mutagenesis and activity assay

DGAT1 mutants were generated by the QuickChange Site-Directed Mutagenesis kit (Agilent) by using the protocol provided by the manufacturer. The desired mutations were confirmed by DNA sequencing. Plasmid was amplified in Stbl2 competent cells (Thermo Fisher Scientific) before transfection. For the acyl-transferase activity, DGAT1 activity was measured in vitro as published⁴². In brief, *DGAT1*^{-/-} SUM159 cells, generated by the CRISPR–Cas9 system, were transfected with constructs to express wild type or mutant DGAT1 with the FuGENE HD transfection reagent (Promega) per the manufacturer's instructions. Cells were collected three days after transfection, and the cell pellet was washed with PBS before being snap-frozen in liquid nitrogen and stored in -80 °C. The cell pellet was lysed by sonication in ice-cold lysis buffer containing 250 mM sucrose, 50 mM Tris-HCl (pH 7.4) and EDTA-free protease inhibitor cocktail (Sigma-Aldrich). The unbroken debris was removed by centrifugation at 5,000 rpm, 4 °C for 5 min on a bench-top centrifuge. Each 100 µl of cell lysate (equivalent to ~5 × 10⁵ cells) was incubated with 25 µM DGAT2 inhibitor PF-06424439 (Sigma-Aldrich) on ice for 30 min. Then, 100 µl of reaction mixture was added to the cell lysate to a final concentrations of 100 mM Tris-HCl (pH 7.4), 25 mM MgCl₂, 0.625 g l⁻¹ delipidated BSA, 200 µM 1,2-dioleoylglycerol and 50 µM oleoyl-CoA containing 0.2 µCi [¹⁴C]-oleoyl-CoA as a tracer (American Radiolabelled Chemicals). In some experiments, DGAT1 inhibitor PF-04620110 (Sigma-Aldrich) was added at final concentration of 10 µM into the reaction mixture. For activity assays comparing different acyl-CoA substrates, 0.2 µCi [¹⁴C]-1,2 diacylglycerol was spiked into 200 µM cold DAG substrate, and 50 µM acyl-CoA with different acyl-CoA substrate was used for individual reactions. Reactions were carried out at 37 °C with gentle agitation for 30 min or as indicated. The reactions were quenched by adding chloroform/methanol (2:1 v/v), followed by 2% phosphoric acid for phase separation. The organic phase was collected, dried, resuspended in chloroform and loaded on a silica gel thin-layer chromatography (TLC) plate (Analtech). Lipids were

resolved in a solvent system consisting of hexane, diethyl ether and acetic acid (80:20:1 v/v/v). The radioactivity of the triacylglycerol bands was revealed by Typhoon FLA 7000 phosphor imager (GE Healthcare Life Sciences) and quantified by Quantity One (V4.6.6). The final activity was normalized to the DGAT1 protein amount in the cell lysate that was measured by immunoblotting against DGAT1 or the GFP tag fused at DGAT1 N-terminal end⁴². A known amount of purified DGAT1 served as the control for quantification of DGAT1 from cell lysate.

Reporting summary

Further information on research design is available in the Nature Research Reporting Summary linked to this paper.

Data availability

The 3D cryo-EM density maps have been deposited into the Electron Microscopy Data Bank under accession numbers EMD-21461 (oleoyl-CoA free), EMD-21481 (with intact oleoyl-CoA) and EMD-21488 (with cleaved oleoyl-CoA). The coordinates have been deposited into the Protein Data Bank (PDB) with accession numbers 6VYI (oleoyl-CoA free) and 6VZI (with intact oleoyl-CoA). All other undeposited models and data are available upon request.

- Booth, D. S., Avila-Sakar, A. & Cheng, Y. Visualizing proteins and macromolecular complexes by negative stain EM: from grid preparation to image acquisition. *J. Vis. Exp.* **58**, e3227 (2011).
- Mastroratte, D. N. Automated electron microscope tomography using robust prediction of specimen movements. *J. Struct. Biol.* **152**, 36–51 (2005).
- Liao, M., Cao, E., Julius, D. & Cheng, Y. Single particle electron cryo-microscopy of a mammalian ion channel. *Curr. Opin. Struct. Biol.* **27**, 1–7 (2014).
- Zheng, S. Q. et al. MotionCor2: anisotropic correction of beam-induced motion for improved cryo-electron microscopy. *Nat. Methods* **14**, 331–332 (2017).
- Rohou, A. & Grigorieff, N. CTFFIND4: Fast and accurate defocus estimation from electron micrographs. *J. Struct. Biol.* **192**, 216–221 (2015).
- Ru, H. et al. Molecular mechanism of V(D)J recombination from synaptic RAG1–RAG2 complex structures. *Cell* **163**, 1138–1152 (2015).
- Scheres, S. H. A Bayesian view on cryo-EM structure determination. *J. Mol. Biol.* **415**, 406–418 (2012).
- Scheres, S. H. RELION: implementation of a Bayesian approach to cryo-EM structure determination. *J. Struct. Biol.* **180**, 519–530 (2012).
- Kucukelbir, A., Sigworth, F. J. & Tagare, H. D. Quantifying the local resolution of cryo-EM density maps. *Nat. Methods* **11**, 63–65 (2014).
- Adams, P. D. et al. PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr. D* **66**, 213–221 (2010).
- Emsley, P. & Cowtan, K. Coot: model-building tools for molecular graphics. *Acta Crystallogr. D* **60**, 2126–2132 (2004).
- Pettersen, E. F. et al. UCSF Chimera—a visualization system for exploratory research and analysis. *J. Comput. Chem.* **25**, 1605–1612 (2004).
- Chen, V. B. et al. MolProbity: all-atom structure validation for macromolecular crystallography. *Acta Crystallogr. D* **66**, 12–21 (2010).
- Wallace, A. C., Laskowski, R. A. & Thornton, J. M. LIGPLOT: a program to generate schematic diagrams of protein–ligand interactions. *Protein Eng.* **8**, 127–134 (1995).
- Ran, F. A. et al. Genome engineering using the CRISPR–Cas9 system. *Nat. Protoc.* **8**, 2281–2308 (2013).
- Gluchowski, N. L. et al. Identification and characterization of a novel DGAT1 missense mutation associated with congenital diarrhea. *J. Lipid Res.* **58**, 1230–1237 (2017).
- Pei, J., Kim, B. H. & Grishin, N. V. PROMALS3D: a tool for multiple protein sequence and structure alignments. *Nucleic Acids Res.* **36**, 2295–2300 (2008).
- Robert, X. & Gouet, P. Deciphering key features in protein structures with the new ENDscript server. *Nucleic Acids Res.* **42**, W320–W324 (2014).
- Notredame, C., Higgins, D. G. & Heringa, J. T-Coffee: A novel method for fast and accurate multiple sequence alignment. *J. Mol. Biol.* **302**, 205–217 (2000).

Acknowledgements We thank Z. Li, S. Sterling, R. Walsh and S. Rawson at the Harvard cryo-EM centre and C. Xu, K. Song and K. Lee at the University of Massachusetts cryo-EM facility for electron microscopy data collection; members of the Liao and Farese and Walther laboratories for helpful discussions; B. Schulman and T. Rapoport for comments on the manuscript; and G. Howard for editorial assistance. This work was supported by National Institutes of Health grants R01GM124348 (to R.V.F.), and R01GM097194 (to T.C.W.). X.S. was supported by the American Heart Association postdoctoral fellowship 18POST34030308. T.C.W. is an investigator of the Howard Hughes Medical Institute.

Author contributions X.S., M.L., T.C.W. and R.V.F. conceived the project. X.S. performed protein expression, purification and reconstitution into amphipol, prepared cryo-EM grids, processed cryo-EM data and built the atomic models. M.L. advised on cryo-EM data processing. K.W., N.L.G. and X.S. performed the mutagenesis and activity studies. S.D.E. made genetically modified cell lines. X.S., M.L., T.C.W. and R.V.F. wrote the manuscript. All authors analysed and discussed the results and contributed to the manuscript.

Competing interests The authors declare no competing interests.

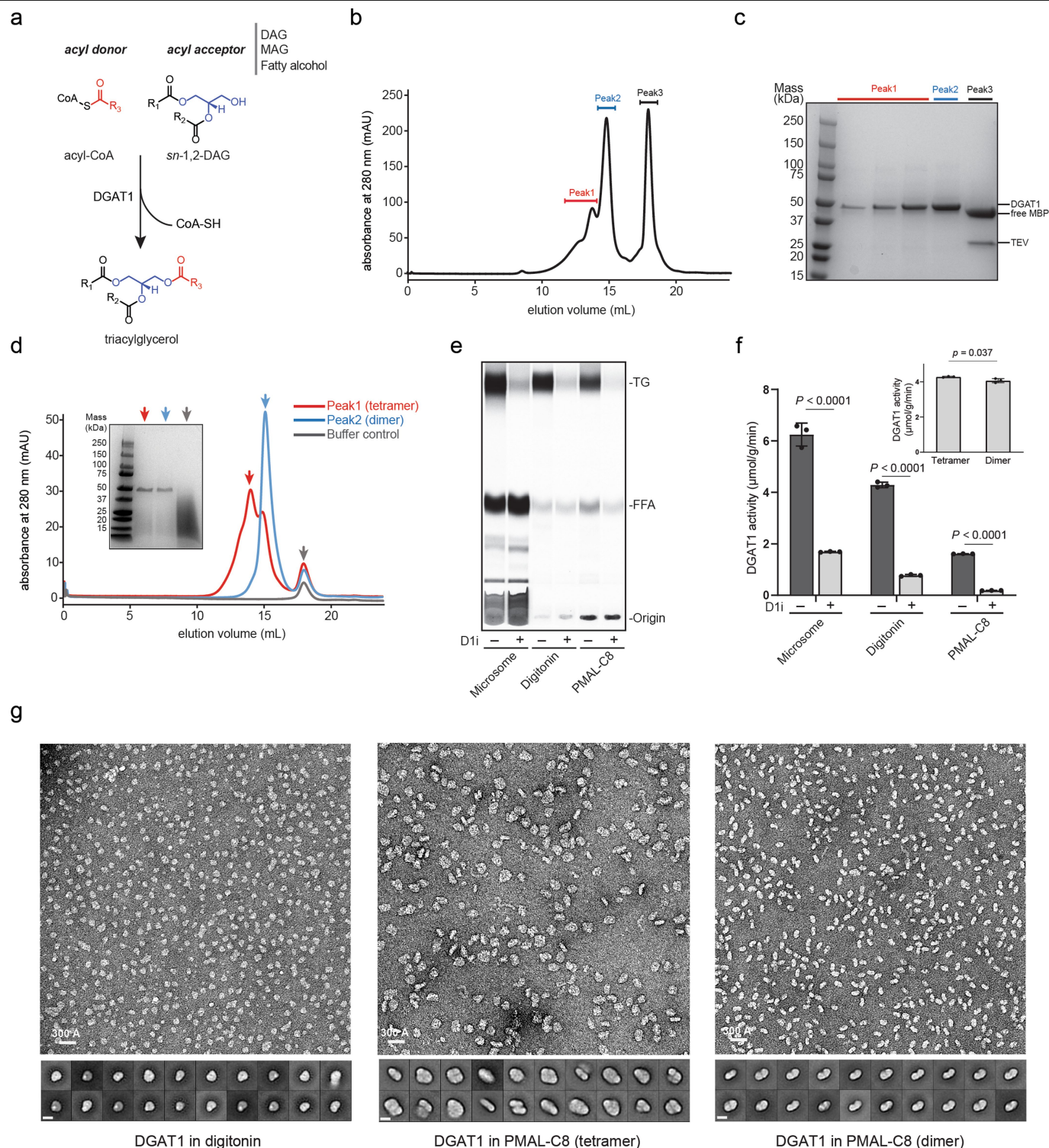
Additional information

Supplementary information is available for this paper at <https://doi.org/10.1038/s41586-020-2289-6>.

Correspondence and requests for materials should be addressed to M.L., T.C.W. or R.V.F.

Peer review information *Nature* thanks David Drew, Savvas N. Savvides and the other, anonymous, reviewer(s) for their contribution to the peer review of this work.

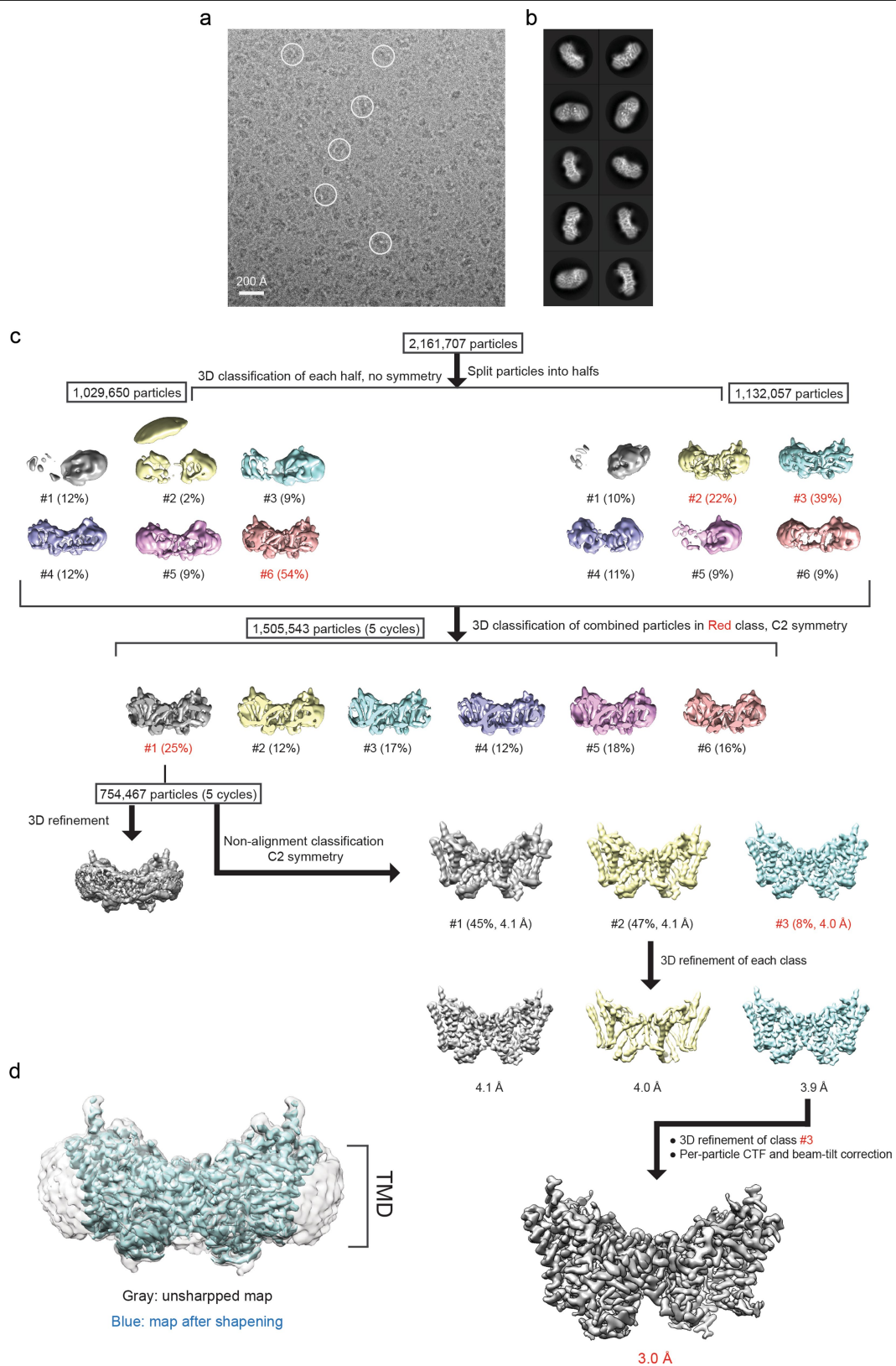
Reprints and permissions information is available at <http://www.nature.com/reprints>.



Extended Data Fig. 1 | Purification and characterization of human DGAT1.

a The acyl-transfer reaction catalysed by DGAT1. The enzyme utilizes acyl-CoA as the sole acyl donor and recognizes different lipid molecules (that is, DAG, monoacylglycerol (MAG) and fatty alcohol) as the acyl acceptor. The panel shows the reaction with *sn*-1,2-DAG as the acyl-acceptor and triacylglycerol as the product. The acyl-group in acyl-CoA is coloured in red, and the glycerol backbone in DAG is coloured in blue. **b, c**, Gel-filtration profile (b) and SDS-PAGE analysis (c) of purified DGAT1 digitonin. Peaks 1 and 2 containing purified DGAT1 by SDS-PAGE analysis were separately collected and pooled. **d**, Gel-filtration of DGAT1 reconstituted in the amphipol PMAL-C8 that was purified from b. A red (tetramer) and blue (dimer) arrow denote the different oligomerization states of DGAT1. The SDS-PAGE analysis of each peak is shown in the insert. **e**, Activity analysis of

DGAT1-overexpressed microsome or purified enzyme with or without DGAT1 inhibitor (D1i). The protein from the digitonin sample of peak 2 in **b** and PMAL-C8 sample of the dimer peak in **d** were used for the assay. The reaction product of triacylglycerol was separated and analysed by TLC. FFA, free fatty acid. **f**, Quantification of triacylglycerol product shown in **e** by phosphorimaging. The insert shows the activity of DGAT1 tetramer and dimer from **d** (mean \pm s.d., $n = 3$ independent experiments). Analysis was performed using two-way ANOVA with Sidak's post hoc test. **g**, Representative negative-stain electron micrograph and 2D averages of purified DGAT1 in digitonin (peak 2 in **b**), and the tetramer and dimer species of DGAT1 in PMAL-C8 (red and blue peaks in **d**, respectively). The bar in 2D average is 100 Å. Experiments shown in **b–e** and **g** were repeated at least three times with similar results. TG, triglyceride.

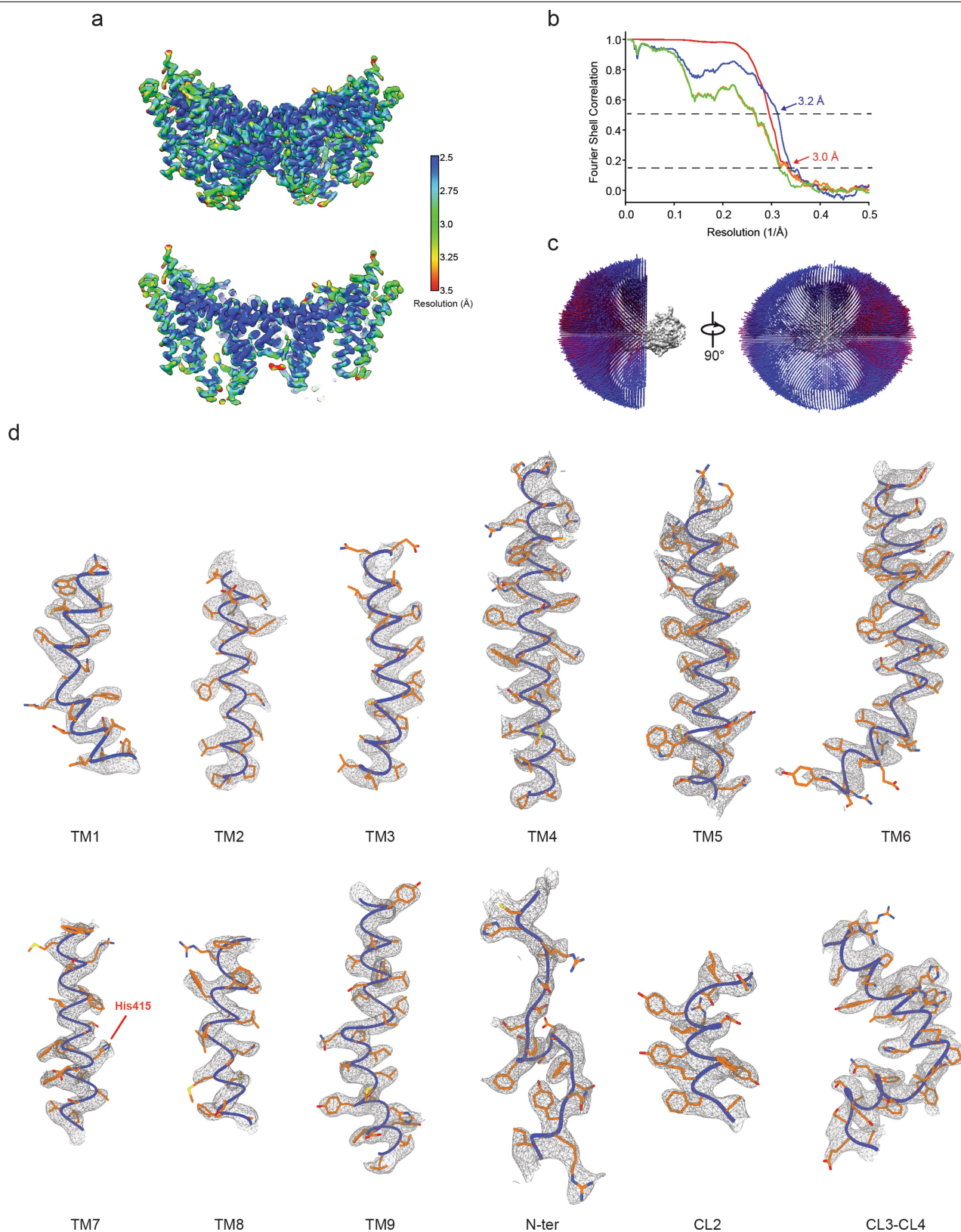


Extended Data Fig. 2 | See next page for caption.

Article

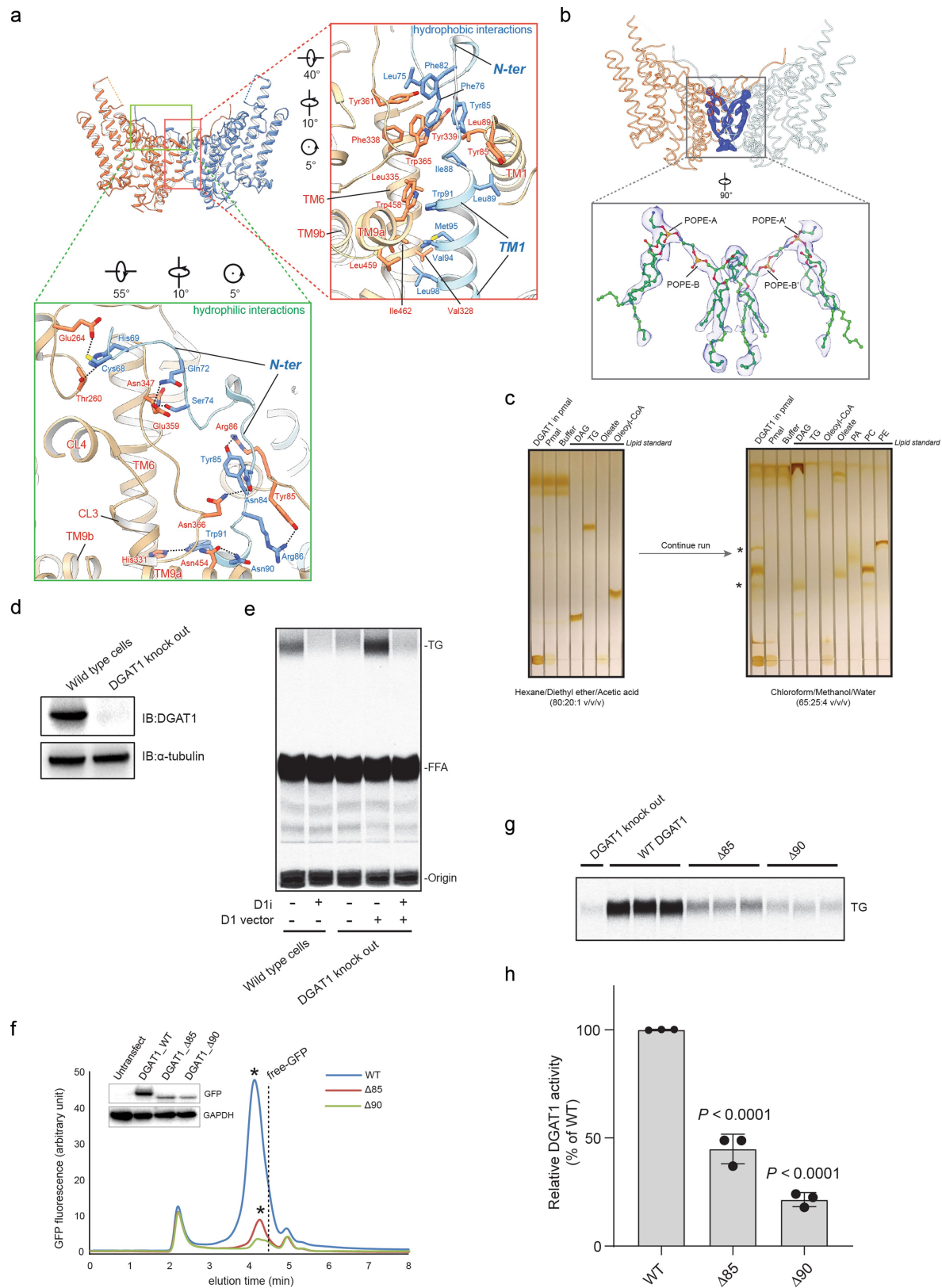
Extended Data Fig. 2 | Cryo-EM image processing of human DGAT1 in PMAL-C8. **a**, Representative cryo-EM image of DGAT1 in PMAL-C8. Some DGAT1 particles are outlined by circles. **b**, The 2D class averages of cryo-EM particle images. The box size of 2D averages is 210 Å. **c**, Three-dimensional classification and refinement of cryo-EM particles. The initial particle stack was split into two stacks due to the large number of particles for 3D classification. After the first round of classification without imposing symmetry, all of the particles classified into one best class (class 1 contains the most abundant 25% particles) in the final five iterations (indicated as '5 cycles')

were kept for further processing. This subset of particles was further classified into three classes by non-alignment classification with C2 symmetry. Afterwards, another round of refinement was performed on each individual class. Among them, class 3 exhibited the highest estimated resolution by Relion and the best side-chain signals by visual inspection and was kept for per-particle contrast transfer function (CTF) and beam-tilt corrections. The resulting cryo-EM map was used for the final cryo-EM maps, contoured at 5σ . **d**, Unsharpened map (grey) superimposed with the final DGAT1 cryo-EM map (blue) showing the detergent micelle-like signals around DGAT1.



Extended Data Fig. 3 | Single-particle cryo-EM analysis of DGAT1 reconstituted in PMAL-C8. a, Local resolution of the final cryo-EM map of DGAT1. A sliced view of local resolution is shown in the lower panel. **b,** FSC curves: gold-standard FSC curve between the two half maps with indicated resolution at FSC = 0.143 (red); FSC curve between the atomic model and the final map with indicated resolution at FSC = 0.5 (blue); FSC curve between half

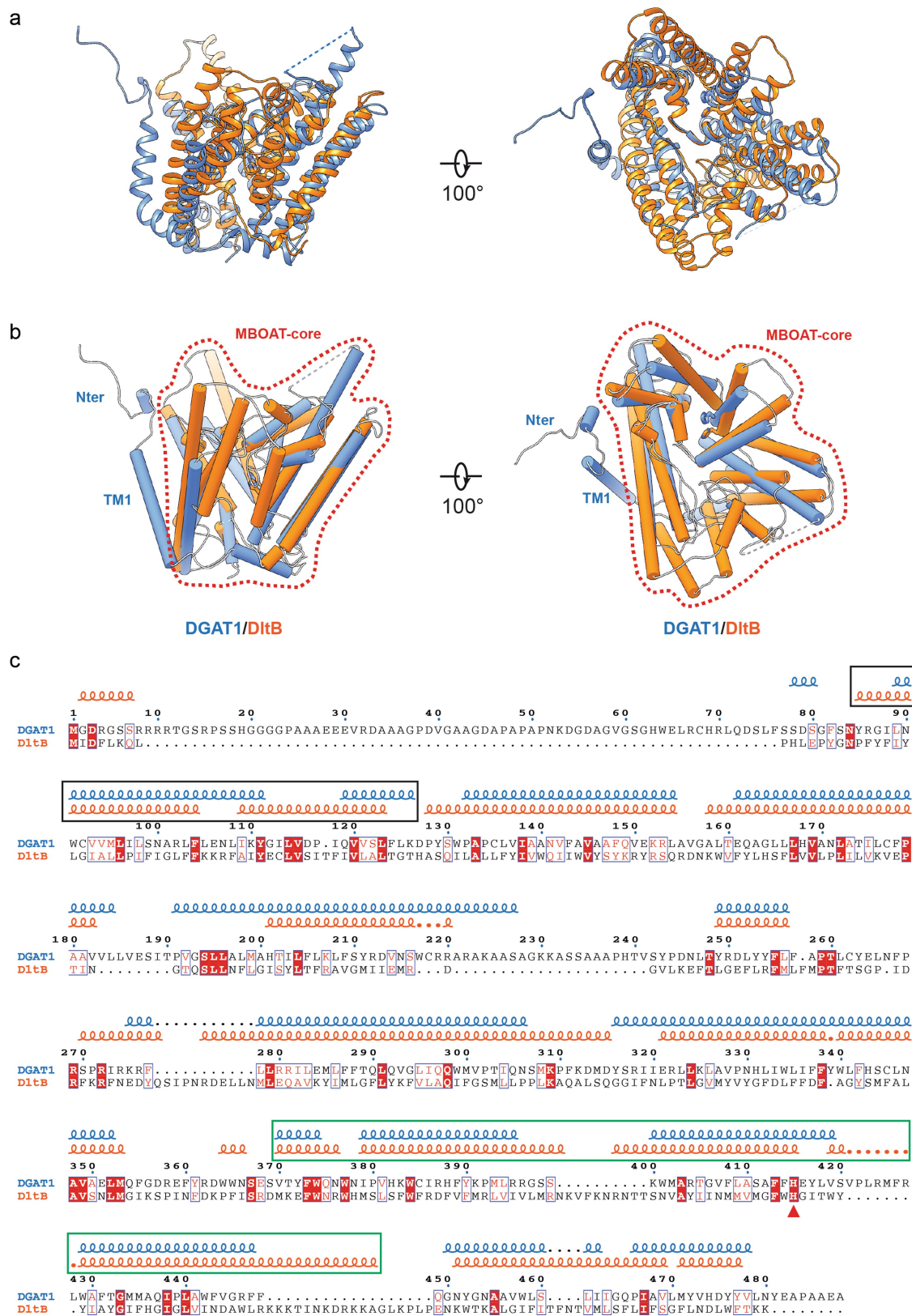
map 1 (orange) or half map 2 (green) and the atomic model refined against half map 1. **c,** Cutaway views of angular distribution of particle images included in the final 3D reconstruction. **d,** Cryo-EM densities superimposed with the atomic model for individual transmembrane helices (TM1-TM9), resolved N-terminal region (N-ter), and helices in the cytosolic loop region (CL2-CL4). The conserved His415 residue is also labelled. Maps are contoured at 4σ.



Extended Data Fig. 4 | See next page for caption.

Extended Data Fig. 4 | Dimer assembly of DGAT1 and phospholipid molecules residing at the dimer interface. **a**, Structural details of the DGAT1 dimer interaction. The resolved N-terminal domain (N-ter, amino acids 68–86) of DGAT1 forms a hydrogen-bond network with the opposing subunit depicted in a green box. Dashed lines denote hydrogen-bonding pairs of residues. Hydrophobic interactions mediating the DGAT1 dimer are shown in a red box. **b**, Densities attributed to phospholipid lipids at the DGAT1 dimer interface are shown as blue surface. The lower panel shows four 1-palmitoyl-2-oleoyl-*sn*-glycero-3-phosphoethanolamine (POPE) molecules modelled into the density; each pair of POPE densities was symmetry-related and labelled as POPE-A/B and POPE-A'/B'. Maps are contoured at 2.5 σ . **c**, TLC analysis of lipids that co-purified with DGAT1 by iodine staining. The asterisks indicate the presence of phosphatidylethanolamine (PE) and DAG in purified DGAT1. PA, phosphatidic acid; PC, phosphatidylcholine. **d**, **e**, The SUM159

DGAT1-knockout cell line system analysed by western blot (**d**) and DGAT1 activity analysis by using lysates from SUM159 DGAT1-knockout cells or cells transiently overexpressing DGAT1 (**e**). **f**, Truncation of N termini reduces DGAT1 expression. Fluorescence size-exclusion chromatography (FSEC) and western blot analyses of two N-terminal truncations lacking the first 85 and 90 residues. The asterisks in the gel-filtration profiles denote the DGAT1-containing peak. The dashed line marks the peaking containing free GFP. **g**, **h**, Truncation of N-terminal region reduces DGAT1 activity. TLC analysis of triacylglycerol products is shown in **g**. Each assay was performed in triplicate. The final triacylglycerol products were normalized to the protein expression level with the quantification results shown in **h**. Mean \pm s.d., $n = 3$ independent experiments. Analysis was performed using one-way ANOVA with Dunnett's post hoc test. Analysis shown in **c** was performed once; experiments in **d–g** were repeated three times with similar results.

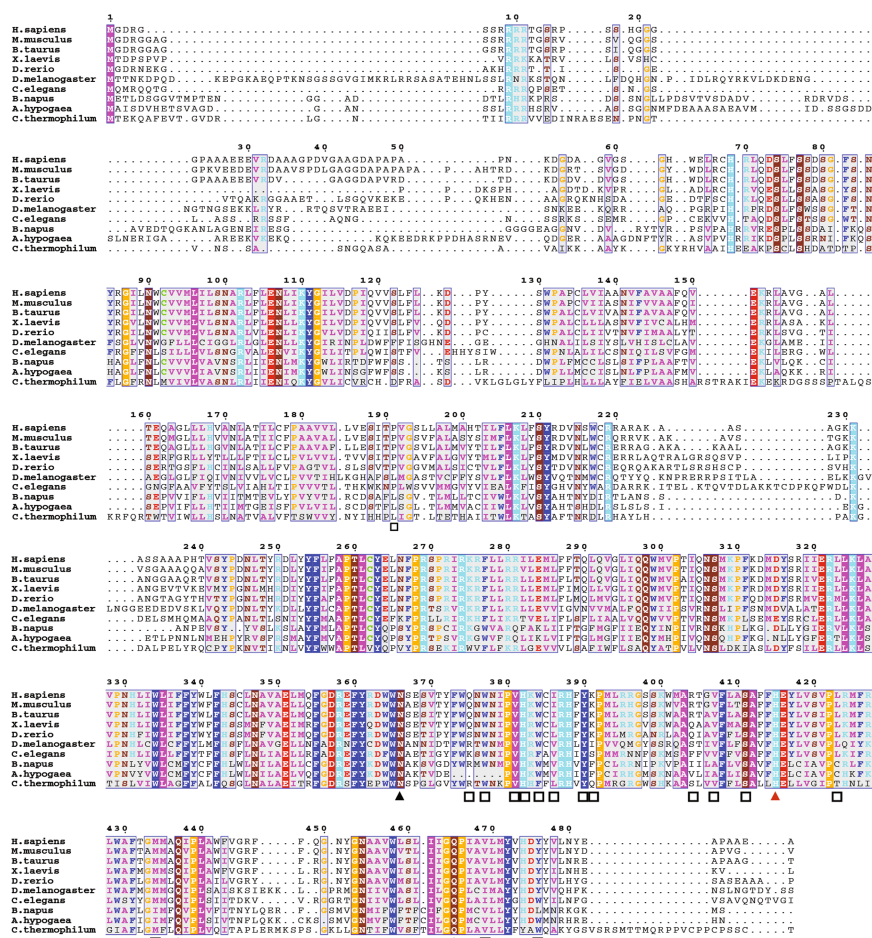


Extended Data Fig. 5 | See next page for caption.

Extended Data Fig. 5 | Structural comparison of DGAT1 with DltB. a, Structural superposition of DGAT1 (blue) and DltB (orange, PDB ID: 6BU1)¹⁷. The two structures were superimposed with an r.m.s.d. of 1.39 Å over 360 matched Cα positions. **b,** As **a**, but the protein structures are shown as cylinders with the conserved histidine residue shown as sticks. The area denoted by the red dashed line exhibits a similar overall architecture in DGAT1 and DltB (MBOAT core) that is not found in other protein structures. Note that beyond this common area, the resolved N-terminal and TM1 regions in DGAT1 appear as extra structural elements beyond the MBOAT-core region as compared to DltB.

c, Structure-based sequence alignment of DGAT1 and DltB. The squiggles on the top represent α-helices in DGAT1 (blue) and DltB (orange). Sequences in the black rectangle indicate the most structurally variable region in the two enzymes. In DGAT1, these regions are involved in dimer formation, whereas in DltB, the equivalent two helices (TM1 and TM2) seal off the lateral cavity within the membrane (see main text). Sequence in green rectangle denotes the alanyl-donor binding pocket in DltB. The red triangle denotes the conserved histidine residue. The alignment was performed with PROMALS3D⁴³, and the final alignment figure was generated with ESPript 3.0⁴⁴.

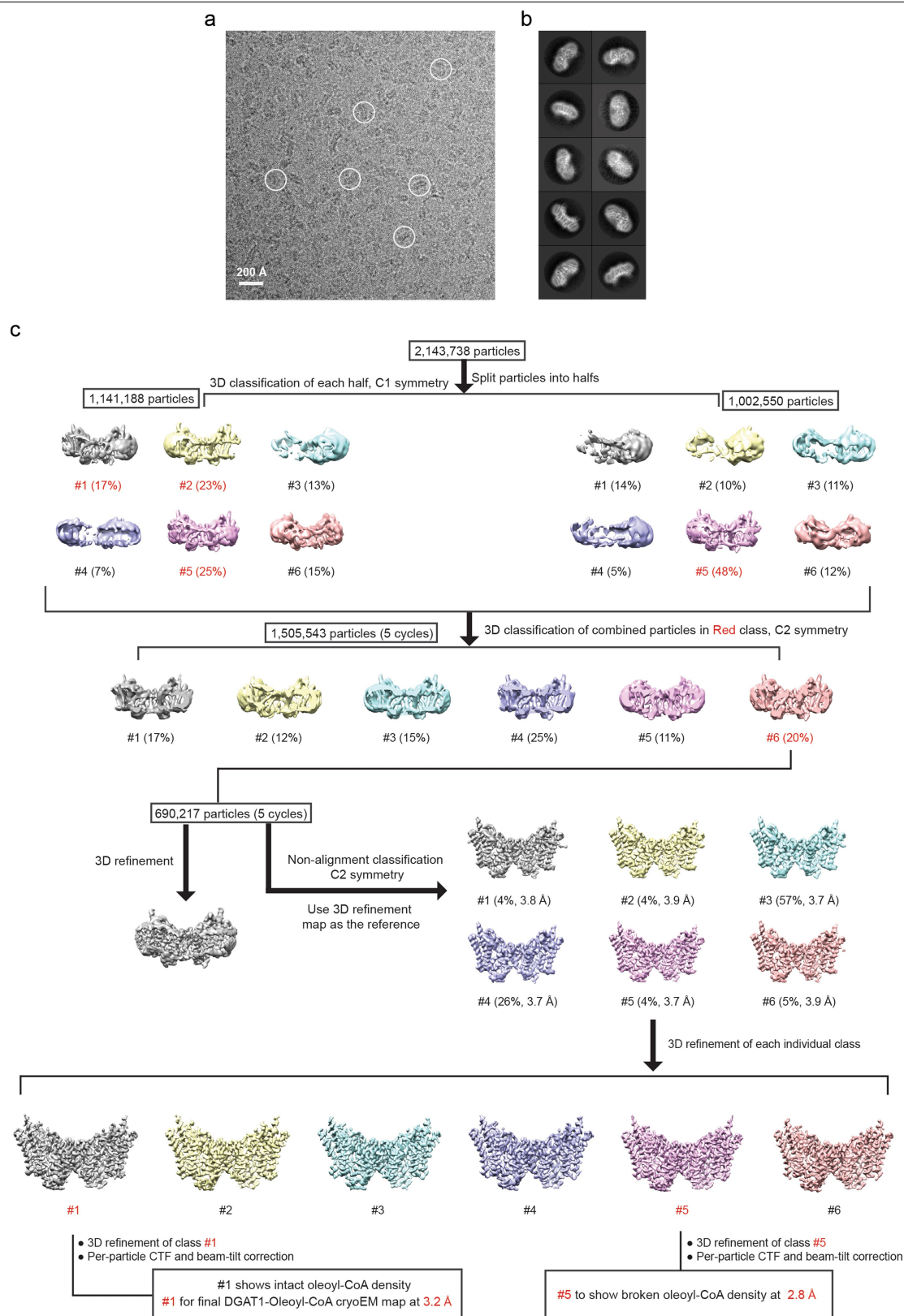
a



Extended Data Fig. 6 | DGAT1 sequence alignment and conservation

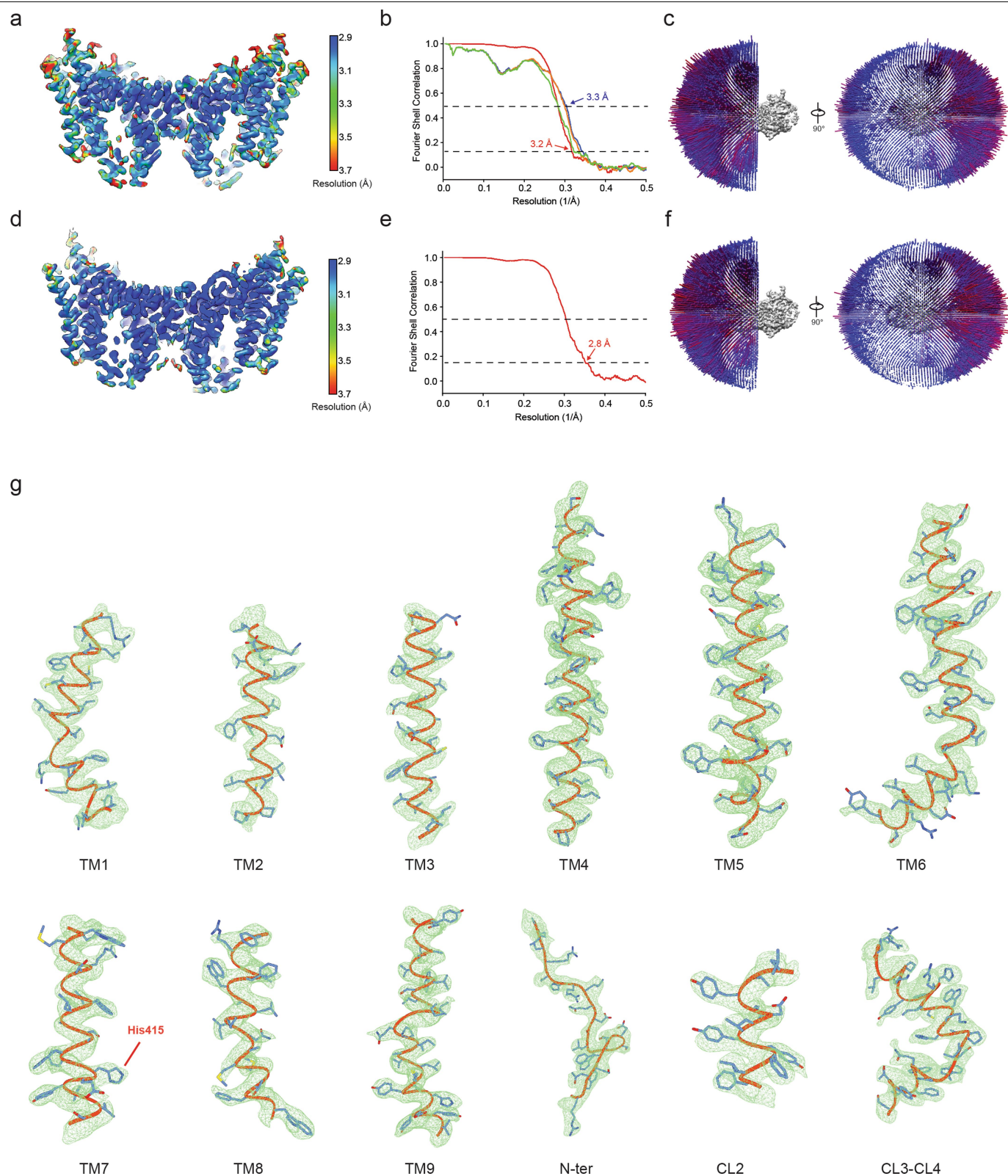
analysis of MBOAT enzymes. a. Sequence alignment of DGAT1 enzymes from *Homo sapiens*, *Mus musculus*, *Bos taurus*, *Xenopus laevis*, *Danio rerio*, *Drosophila melanogaster*, *Caenorhabditis elegans*, *Brassica napus*, *Arachis hypogaea* and *Chaetomium thermophilum*. The colour scheme of amino acids is based on their physicochemical properties. A red triangle denotes the highly conserved histidine residue; black triangles denote highly conserved polar residues in the DGAT1 active centre; black squares denote residues interacting with oleoyl-CoA. The residue numbers for human DGAT1 are indicated at the top. The sequence alignment was performed with T-Coffee⁴⁵, and the final

alignment figure was generated with ESPrpt 3.0⁴⁴. **b.** Sequence alignment of MBOAT enzymes that acylate lipids (DGAT1 and ACAT1) or proteins (PORCN and GOAT). Structural information of DGAT1 was incorporated into the alignment, where the regions containing a cluster of conserved residues among MBOAT enzymes were labelled as blue boxes. Note the alignment starts at Arg250 of DGAT1. **c.** Mapping the conserved blue region shown in **b** into the DGAT1 structure. The DGAT1 structure is shown as grey cylinders. Blue areas denote the conserved region among MBOATs. Note the acyl-CoA binding tunnel in DGAT1 containing the most conserved region among MBOATs (blue area in **b**) is highlighted by the dashed circle.



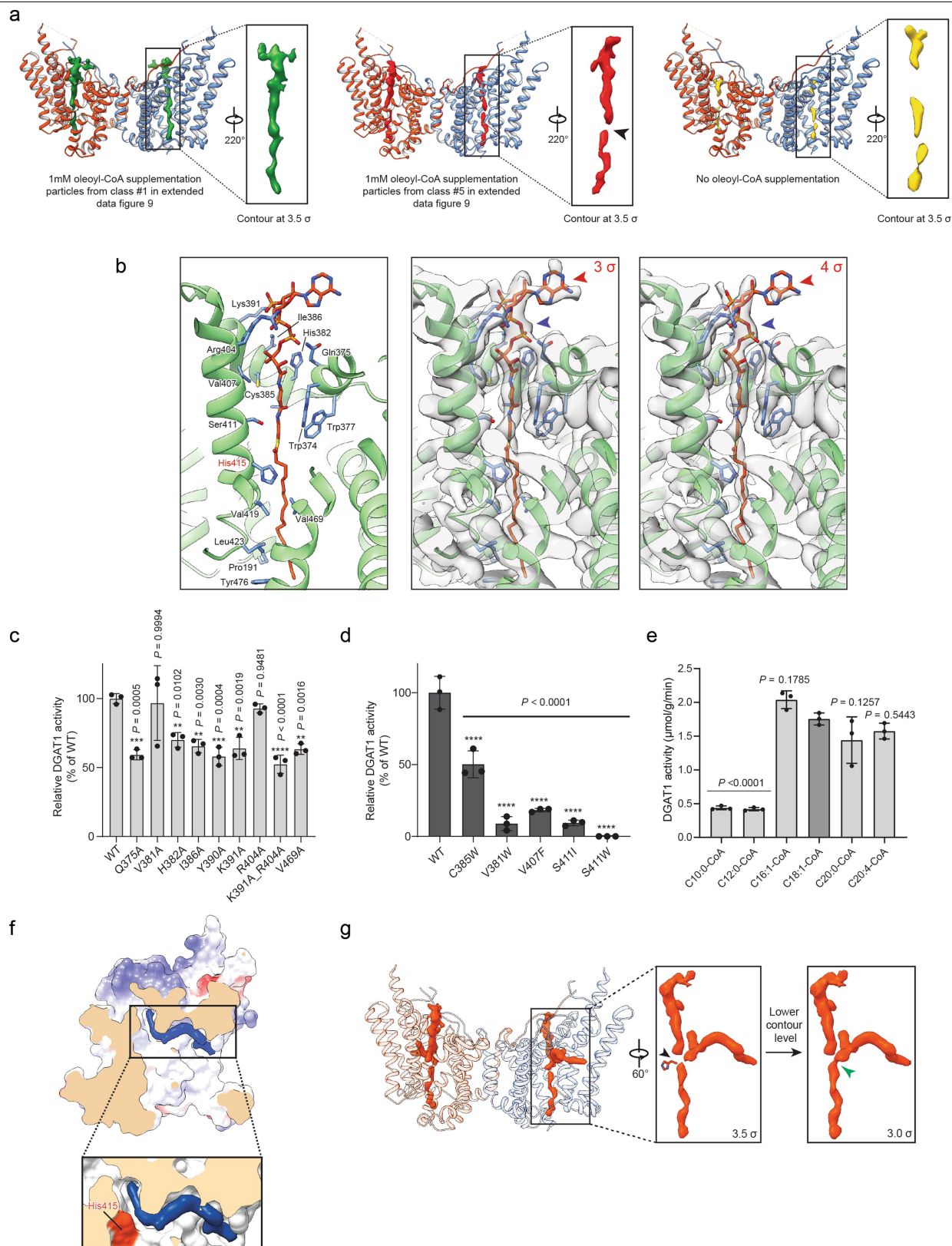
Extended Data Fig. 7 | Cryo-EM image processing of human DGAT1 supplemented with acyl-CoA substrate. **a**, Representative cryo-EM image of DGAT1 with oleoyl-CoA. Some DGAT1 particles are marked by circles. **b**, 2D class averages of cryo-EM particle images with the box size of 210 Å. **c**, Three-dimensional classification and refinement of cryo-EM particle images. The processing flow is similar to the dataset without acyl-CoA treatment (Extended Data Fig. 2). After the second round of 3D classification, all particles

classified into class 6 during the final five iterations (indicated as '5 cycles') were kept for next round of non-alignment classification into six classes. Particles in each class (estimated resolution shown in parentheses) were individually refined. After per-particle CTF and beam-tilt corrections, two resulting maps were chosen to represent DGAT1 complexed with the intact and broken oleoyl-CoA molecule.



Extended Data Fig. 8 | Single-particle cryo-EM analysis of DGAT1 with acyl-CoA substrate. **a**, Local resolution of the final cryo-EM map of DGAT1 (shown as sliced view) complex with an intact oleoyl-CoA molecule. **b**, FSC curves: gold-standard FSC curve between the two half maps with indicated resolution at FSC = 0.143 (red); FSC curve between the atomic model and the final map with indicated resolution at FSC = 0.5 (blue); FSC curve between half map 1 (orange) or half map 2 (green) and the atomic model refined against half

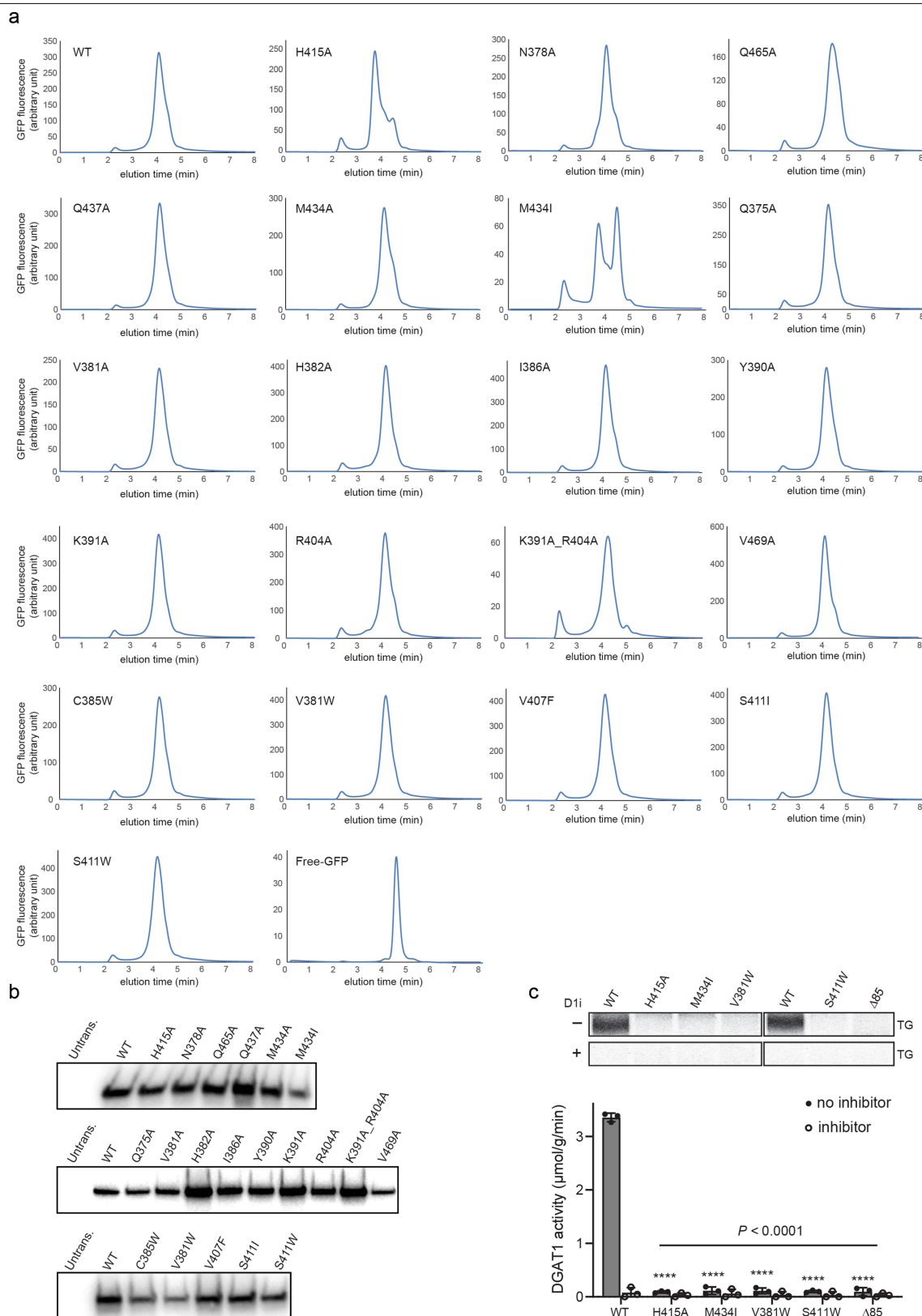
map. **c**, Cutaway views of angular distribution of particle images included in the final 3D reconstruction. **d-f**, Similar to **a-c**, but showing the cryo-EM analysis of the map with a broken oleoyl-CoA density. Note that in **e**, only the FSC curve between half maps was calculated. **g**, Cryo-EM densities with the intact oleoyl-CoA density (shown in **a-c**) superimposed with the atomic model for individual transmembrane helices similar to that shown in Extended Data Fig. 3d. Maps are contoured at 4 σ .



Extended Data Fig. 9 | See next page for caption.

Extended Data Fig. 9 | Cryo-EM density analysis of the bound acyl-CoA and lipid-like density and mutagenesis study of residues interacting with oleoyl-CoA substrate. **a**, A side-by-side density comparison of the bound oleoyl-CoA molecule, oleoyl-CoA with a broken density, and density in the acyl-CoA binding pocket without oleoyl-CoA treatment. **b**, Cryo-EM density map of the oleoyl-CoA molecule and the surrounding region. Maps are shown at different contour levels to illustrate mobility of the adenosine moiety (red arrowhead) of oleoyl-CoA. By contrast, the phosphopantetheine region (green arrowhead) exhibits a stronger density signal. **c**, Activities of acyl-CoA binding tunnel alanine mutants of DGAT1 expressed in DGAT1-knockout cells. **d**, Activities of the bulky side-chain substitutions of residues in the acyl-CoA binding tunnel. Activities in **c** and **d** were normalized to the amount of DGAT1 protein expressed (mean \pm s.d., $n=3$ independent experiments). Analysis was

performed using one-way ANOVA with Dunnett's post hoc test. **e**, Effect of the acyl chain of the acyl-donor substrate on DGAT1 activity. The analyses used purified DGAT1 in amphipol (mean \pm s.d., $n=3$ independent experiments). Analysis was performed using one-way ANOVA with Dunnett's post hoc test comparing C18:1-CoA to other acyl-CoA substrates (coloured in dark grey). **f**, The zoomed-in views of the lipid-like density in the oleoyl-CoA-free DGAT1 monomer structure shown as surface representation. The protein surface at His415 in the zoomed-in view is coloured in orange. The map is contoured at 3.5σ . **g**, Cryo-EM density in orange showing both the bound acyl-CoA with a broken signal and the lipid-like density. A black arrowhead denotes the disconnected acyl chain in oleoyl-CoA molecule. A cyan arrowhead indicates the connected density between the acyl chain of oleoyl-CoA and the uncharacterized lipid-like density at a lower contour level.



Extended Data Fig. 10 | FSEC, western blot and activity analyses of DGAT1 mutants. **a**, FSEC profiles of wild type and all tested mutants in this study. Wild-type DGAT1 and mutants were transiently expressed as GFP-fusion proteins in HEK293 cells. The folding of each mutant was analysed by size-exclusion chromatography monitoring GFP fluorescence. All tested mutants exhibit peaks similar to the wild-type DGAT1 protein, suggesting their correct overall protein folding. **b**, Western blot analyses of wild-type and tested

DGAT1 mutant proteins expressed in SUM159 DGAT1-knockout cells. Untrans., untransfected control. **c**, Activity of wild type and selected DGAT1 mutants purified by His-tag affinity chromatography purification. The top panel of the raw TLC plates shows the formation of triacylglycerol. The assays were performed with or without DGAT1 inhibitor (D1i) (mean \pm s.d., $n = 3$ independent experiments). One-way ANOVA with Dunnett's post hoc test was applied. Experiments shown in **a–c** were repeated twice with similar results.

Extended Data Table 1 | Cryo-EM data collection, refinement and validation statistics

	DGAT1 in PMAL-C8 (EMD-21461) (PDB 6VYI)	DGAT1 with intact oleoyl-CoA (EMDB-21481) (PDB 6VZ1)	DGAT1 with broken oleoyl-CoA density (EMDB-21488)
Data collection and processing			
Magnification	105,000	105,000	105,000
Voltage (kV)	300	300	300
Electron exposure (e-/Å ²)	52	43	43
Defocus range (µm)	1.5-3.0	1.0 - 2.5	1.0 - 2.5
Pixel size (Å)	0.825	0.83	0.83
Symmetry imposed	C2	C2	C2
Initial particle images (no.)	2,161,707	2,143,738	2,143,738
Final particle images (no.)	61,608	28,165	27,750
Map resolution (Å)	3.0	3.2	2.8
FSC threshold	0.143	0.143	0.143
Map resolution range (Å)	211.2 - 3.0	212.5 - 3.2	212.5 - 2.8
Refinement			
Initial model used (PDB code)	NA	NA	
Model resolution (Å)			
FSC threshold	0.143	0.143	
Model resolution range (Å)	211.2 - 3.0	212.5 - 3.2	
Map sharpening <i>B</i> factor (Å ²)	-120	-103	
Model composition			
Non-hydrogen atoms	6916	7112	
Protein residues	812	822	
Ligands	4	6	
<i>B</i> factors (Å ²)			
Protein	53.8	67.4	
Ligand	44.4	67.7	
R.m.s. deviations			
Bond lengths (Å)	0.01	0.01	
Bond angles (°)	0.74	0.85	
Validation			
MolProbity score	2.1	2.1	
Clashscore	9.3	9.9	
Poor rotamers (%)	0	0	
Ramachandran plot			
Favored (%)	88.8	87.9	
Allowed (%)	11.2	12.1	
Disallowed (%)	0	0	

Reporting Summary

Nature Research wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Research policies, see [Authors & Referees](#) and the [Editorial Policy Checklist](#).

Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

- | n/a | Confirmed |
|-------------------------------------|---|
| <input type="checkbox"/> | <input checked="" type="checkbox"/> The exact sample size (<i>n</i>) for each experimental group/condition, given as a discrete number and unit of measurement |
| <input type="checkbox"/> | <input checked="" type="checkbox"/> A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly |
| <input checked="" type="checkbox"/> | <input type="checkbox"/> The statistical test(s) used AND whether they are one- or two-sided
<i>Only common tests should be described solely by name; describe more complex techniques in the Methods section.</i> |
| <input checked="" type="checkbox"/> | <input type="checkbox"/> A description of all covariates tested |
| <input checked="" type="checkbox"/> | <input type="checkbox"/> A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons |
| <input checked="" type="checkbox"/> | <input type="checkbox"/> A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals) |
| <input checked="" type="checkbox"/> | <input type="checkbox"/> For null hypothesis testing, the test statistic (e.g. <i>F</i> , <i>t</i> , <i>r</i>) with confidence intervals, effect sizes, degrees of freedom and <i>P</i> value noted
<i>Give P values as exact values whenever suitable.</i> |
| <input checked="" type="checkbox"/> | <input type="checkbox"/> For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings |
| <input checked="" type="checkbox"/> | <input type="checkbox"/> For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes |
| <input checked="" type="checkbox"/> | <input type="checkbox"/> Estimates of effect sizes (e.g. Cohen's <i>d</i> , Pearson's <i>r</i>), indicating how they were calculated |

Our web collection on [statistics for biologists](#) contains articles on many of the points above.

Software and code

Policy information about [availability of computer code](#)

Data collection Serial EM version 3.6

Data analysis Phenix (v 1.16); Coot (v 0.8.9); MotionCor2 (v 1.1.0); CTFFIND4 (v 4.1.5); SAMUEL (v 17.05); SamViewer (v 16.01); SPIDER (v 17.05); RELION 3.0; ResMap (v 1.1.4); bfactor (v 1.03); Chimera (v 1.13); LigPlot (v 2.2)

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors/reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Research [guidelines for submitting code & software](#) for further information.

Data

Policy information about [availability of data](#)

All manuscripts must include a [data availability statement](#). This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A list of figures that have associated raw data
- A description of any restrictions on data availability

Three cryoEM maps have been deposited into the Electron Microscopy Data Bank under the accession code EMD-21461, EMD-21481, and EMD-21488; Two protein coordinates have been deposited in the Protein Data Bank with the accession code 6VYI and 6VZ1.

Field-specific reporting

Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

☒ Life sciences ☐ Behavioural & social sciences ☐ Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see [nature.com/documents/nr-reporting-summary-flat.pdf](https://www.nature.com/documents/nr-reporting-summary-flat.pdf)

Life sciences study design

All studies must disclose on these points even when the disclosure is negative.

Sample size	No statistical methods were used to predetermine sample size. Sufficient EM data were collected to achieve adequate single-particle EM analysis and 3D cryo-EM reconstructions.
Data exclusions	No data were excluded from analyses.
Replication	Each experiment was repeated at least two times in independent experiments. Experimental findings were reproduced reliably.
Randomization	This is not relevant to our study, because no grouping was needed.
Blinding	Investigators were not blinded to group allocation, because no grouping was needed for this study.

Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

Materials & experimental systems

n/a	Involved in the study
<input checked="" type="checkbox"/>	<input type="checkbox"/> Antibodies
<input type="checkbox"/>	<input checked="" type="checkbox"/> Eukaryotic cell lines
<input checked="" type="checkbox"/>	<input type="checkbox"/> Palaeontology
<input checked="" type="checkbox"/>	<input type="checkbox"/> Animals and other organisms
<input checked="" type="checkbox"/>	<input type="checkbox"/> Human research participants
<input checked="" type="checkbox"/>	<input type="checkbox"/> Clinical data

Methods

n/a	Involved in the study
<input checked="" type="checkbox"/>	<input type="checkbox"/> ChIP-seq
<input checked="" type="checkbox"/>	<input type="checkbox"/> Flow cytometry
<input checked="" type="checkbox"/>	<input type="checkbox"/> MRI-based neuroimaging

Eukaryotic cell lines

Policy information about [cell lines](#)

Cell line source(s)	HEK293S GnTI- cell (ATCC, CRL-3022). SUM159 CRISPR DGAT1 knock out cell (created from ATCC, CRL-2915)
Authentication	No further authentications were performed for this study.
Mycoplasma contamination	No mycoplasma contamination test was performed for this study.
Commonly misidentified lines (See ICLAC register)	None of the cell lines used are listed in the ICLAC database.

Structure and mechanism of human diacylglycerol *O*-acyltransferase 1

<https://doi.org/10.1038/s41586-020-2280-2>

Received: 24 May 2019

Accepted: 17 March 2020

Published online: 13 May 2020



Lie Wang^{1,6}, Hongwu Qian^{2,6}, Yin Nian^{1,4,6}, Yimo Han^{2,5,6}, Zhenning Ren¹, Hanzhi Zhang¹, Liya Hu¹, B. V. Venkataram Prasad¹, Arthur Laganowsky³, Nieng Yan²✉ & Ming Zhou¹✉

Diacylglycerol *O*-acyltransferase 1 (DGAT1) synthesizes triacylglycerides and is required for dietary fat absorption and fat storage in humans¹. DGAT1 belongs to the membrane-bound *O*-acyltransferase (MBOAT) superfamily, members of which are found in all kingdoms of life and are involved in the acylation of lipids and proteins^{2,3}. How human DGAT1 and other mammalian members of the MBOAT family recognize their substrates and catalyse their reactions is unknown. The absence of three-dimensional structures also hampers rational targeting of DGAT1 for therapeutic purposes. Here we present the cryo-electron microscopy structure of human DGAT1 in complex with an oleoyl-CoA substrate. Each DGAT1 protomer has nine transmembrane helices, eight of which form a conserved structural fold that we name the MBOAT fold. The MBOAT fold in DGAT1 forms a hollow chamber in the membrane that encloses highly conserved catalytic residues. The chamber has separate entrances for each of the two substrates, fatty acyl-CoA and diacylglycerol. DGAT1 can exist as either a homodimer or a homotetramer and the two forms have similar enzymatic activity. The N terminus of DGAT1 interacts with the neighbouring protomer and these interactions are required for enzymatic activity.

DGAT1 (EC 2.3.1.20) is an integral membrane protein that synthesizes triacylglycerides from two substrates: diacylglycerol (DAG) and fatty acyl-CoA¹ (Extended Data Fig. 1). In humans, DGAT1 is highly expressed in epithelial cells of the small intestine and its activity is essential for the absorption of dietary fats⁴. DGAT1 is also found in the liver, in which it synthesizes fat for storage^{5,6}, and in female mammary glands, in which it produces fat in the milk⁷. *Dgat1*^{-/-} mice are viable, and show substantially reduced levels of triacylglycerides in all tissues and resistance to obesity when kept on a high-fat diet^{8,9}. These results have generated considerable interest in DGAT1 as a potential target for the treatment of hypertriglyceridaemia and fatty liver disease¹⁰.

DGAT1 belongs to the large MBOAT superfamily (<http://pfam.xfam.org/family/MBOAT>), members of which are found in all kingdoms of life. In mammals, the MBOAT family includes enzymes that modify lipids or proteins, such as acyl-CoA:cholesterol acyltransferase (ACAT)² and protein-serine *O*-palmitoleoyltransferase (PORCUPINE)³. Members of the MBOAT family have a highly conserved histidine residue that is required for their transferase activity, and are predicted to have eight to eleven transmembrane segments^{2,11,12}. The crystal structure of a bacterial MBOAT, DltB, has previously been described¹³ (Extended Data Fig. 1). However, the structure of DltB may not be a suitable model for human DGAT1 because the sequence identity of the two proteins is low (around 20%).

In vitro activity of purified human DGAT1

Full-length human DGAT1 was overexpressed and purified. DGAT1 purified in the detergent lauryl maltose neopentyl glycol (LMNG) exists mainly as a stable dimer that is partially resistant to the denaturing conditions of SDS-PAGE (Extended Data Fig. 2a, Methods). When DGAT1 was purified with a milder detergent, glyco-diosgenin (GDN), a substantially higher fraction of DGAT1 was in the tetrameric form. Both the dimeric and tetrameric forms of human DGAT1 seem stable in terms of their oligomeric state (Extended Data Fig. 2b). DGAT1 from plants and mammals was previously shown to form either a dimer or a tetramer^{11,14,15}; however, it is not clear whether the oligomeric state has an effect on enzymatic functions.

We established an in vitro functional assay to measure the activity of human DGAT1 (Extended Data Fig. 2d–f, Methods). The initial rate of the enzymatic reaction in different concentrations of oleoyl-CoA can be fitted with a Michaelis–Menten equation for both the dimeric and the tetrameric DGAT1 (Extended Data Fig. 2g, Extended Data Table 2). The tetrameric DGAT1 has a slightly higher velocity of enzyme-catalysed reaction at infinite concentration of substrate (V_{\max}) and the V_{\max} values are equivalent to a turnover rate of around one molecule per second for each DGAT1 protomer. Both forms have a similar Michaelis constant (K_m). We also measured the activity of DGAT1 in cell membranes and found that the V_{\max} is about 50% higher than that of DGAT1 in detergent,

¹Verna and Marrs McLean Department of Biochemistry and Molecular Biology, Baylor College of Medicine, Houston, TX, USA. ²Department of Molecular Biology, Princeton University, Princeton, NJ, USA. ³Department of Chemistry, Texas A&M University, College Station, TX, USA. ⁴Present address: Ion Channel Research and Drug Development Center, Kunming Institute of Zoology, Chinese Academy of Sciences, Kunming, China. ⁵Present address: Department of Material Science and Nanoengineering, Rice University, Houston, TX, USA. ⁶These authors contributed equally: Lie Wang, Hongwu Qian, Yin Nian, Yimo Han. ✉e-mail: nyan@princeton.edu; mzhou@bcm.edu

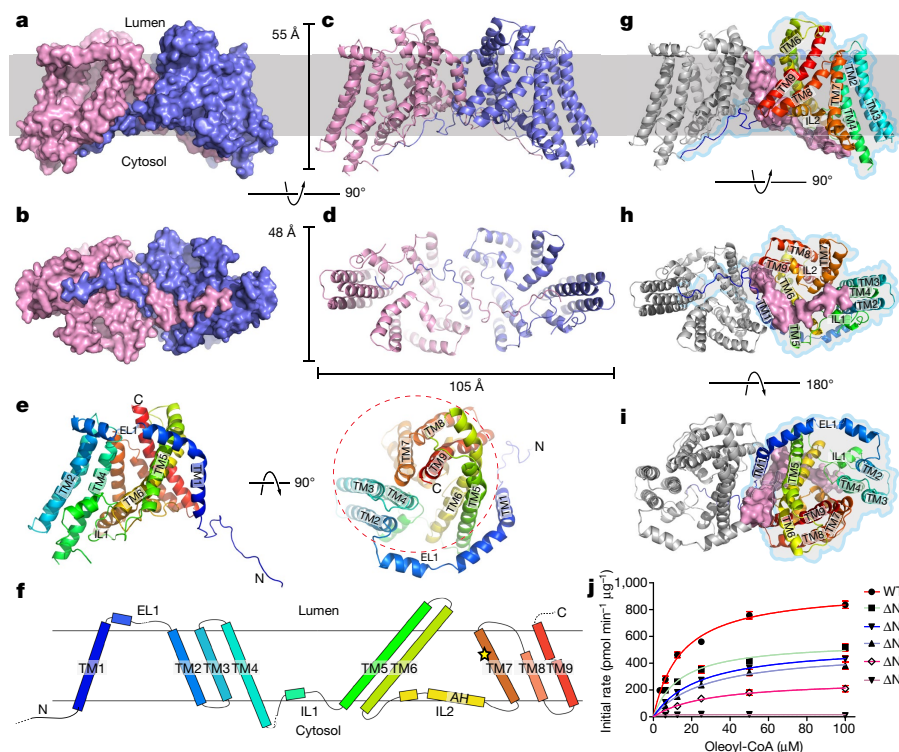


Fig. 1 | Structure of human DGAT1. **a–d**, The structure of the DGAT1 dimer is shown in cartoon and surface representations as viewed from within the plane of the membrane (**a**, **c**), or the intracellular side of the membrane (**b**, **d**). The approximate position of the ER membrane is marked with grey shading. **e**, Cartoon representation of a DGAT1 protomer in two orientations. The MBOAT fold is marked with a dashed circle. **f**, Topology of DGAT1. Unresolved regions in the structure are marked with dashed lines. The position of His415 is

shown with a yellow star. **g–i**, Dimerization interface of DGAT1 viewed in three orientations. One protomer is shown as a grey cartoon but with its TM1 and the N terminus as surface. The other protomer is shown as a rainbow-coloured cartoon and marked with a blue outline. **j**, Enzymatic activity of N-terminal truncations of DGAT1. Data are mean \pm s.e.m. derived from three independent repeats.

whereas the K_m is similar (Extended Data Fig. 2h, Extended Data Table 2). Both the V_{max} and the K_m values reported here are comparable to those previously reported for human DGAT1 in microsomes¹⁶. The enzymatic activity was also measured in different concentrations of DAG for both dimeric and tetrameric DGAT1 and the two forms show similar activity (Extended Data Fig. 2i, Extended Data Table 2). Notably, both datasets were better fit with an allosteric sigmoidal equation (Methods), suggesting that DAG has a regulatory role on DGAT1.

Overall structure of human DGAT1

The structure of human DGAT1 was solved by single-particle cryo-electron microscopy (cryo-EM) (Extended Data Fig. 3a–e, Methods). A density map was reconstructed to an overall resolution of 3.1 Å with an imposed C2 symmetry. The resolution for helices that are close to the core of the dimer reaches 2.7 Å, whereas regions close to the periphery of the dimer have lower resolution, probably owing to their relatively higher mobility (Extended Data Fig. 3f).

The density map is of sufficient quality to allow de novo building of residues 64 to 224 and 239 to 481—which include all the transmembrane helices, one oleoyl-CoA molecule and five partially resolved lipid or detergent molecules—and the structure was refined to proper geometry (Extended Data Fig. 4a–c, Extended Data Table 1). The first 63 residues, last 5 residues and residues 225–238 (which are part of a cytosolic loop) were not resolved. Residues 112 to 120 (part of a luminal loop) were partially resolved and built as poly-alanines.

The DGAT1 dimer is around 105 Å by 55 Å by 48 Å, and is similar in shape to a canoe (Fig. 1a–d). On the basis of the positive-inside rule¹⁷, the N terminus of DGAT1 resides at the cytosolic side (Extended Data Fig. 4d). This assignment is also consistent with the previous consensus from

biochemical studies^{11,18} and allows for unambiguous placement of the C terminus to the luminal side of the endoplasmic reticulum (ER) (Fig. 1e, f). Each DGAT1 protomer has nine transmembrane helices, TM1–TM9, and three long loops: an ER luminal (extracellular) loop EL1 between TM1 and TM2; an intracellular loop IL1 between TM4 and TM5; and a second intracellular loop IL2 between TM6 and TM7. In each protomer, TM2–TM9 and the two intracellular loops IL1 and IL2 form a distinctive structural fold that we name the MBOAT fold (Figs. 1e, f, 2a–d). TM1, which is not part of the MBOAT fold, is isolated and linked to the MBOAT fold by the long ER luminal loop EL1 (residues 110 to 125). EL1 is partially structured and extends around 35 Å along the luminal side of the protein (Fig. 1e, f).

The dimer interface

Although TM1 seems to be suspended in the membrane when a protomer is viewed in isolation, the space between the TM1 and the rest of the protomer (the MBOAT fold) is partially filled by the TM1 from the neighbouring protomer so that the two form a domain-swapped homodimer (Fig. 1b, g–i). Crossover of the TM1 helix allows the N-terminal residues 64–80 of one protomer to interact with both IL1 and IL2 of the neighbouring protomer. TM1 interacts with TM6 and TM9 of the neighbouring subunit, and the two TM1 helices make contact at residues Ser83 and Asn84, which are located close to the intracellular side of the membrane. The rest of the space between the two protomers is filled with two detergent molecules and four partially resolved lipid molecules (Extended Data Figs. 4c, 5a–e). Because these bound detergent and lipid molecules have extensive interactions with DGAT1, they may have important roles in both the structure and the function of DGAT1.

Previous studies on a plant DGAT1 identified part of the N terminus as intrinsically disordered protein, and showed that deletion of the entire

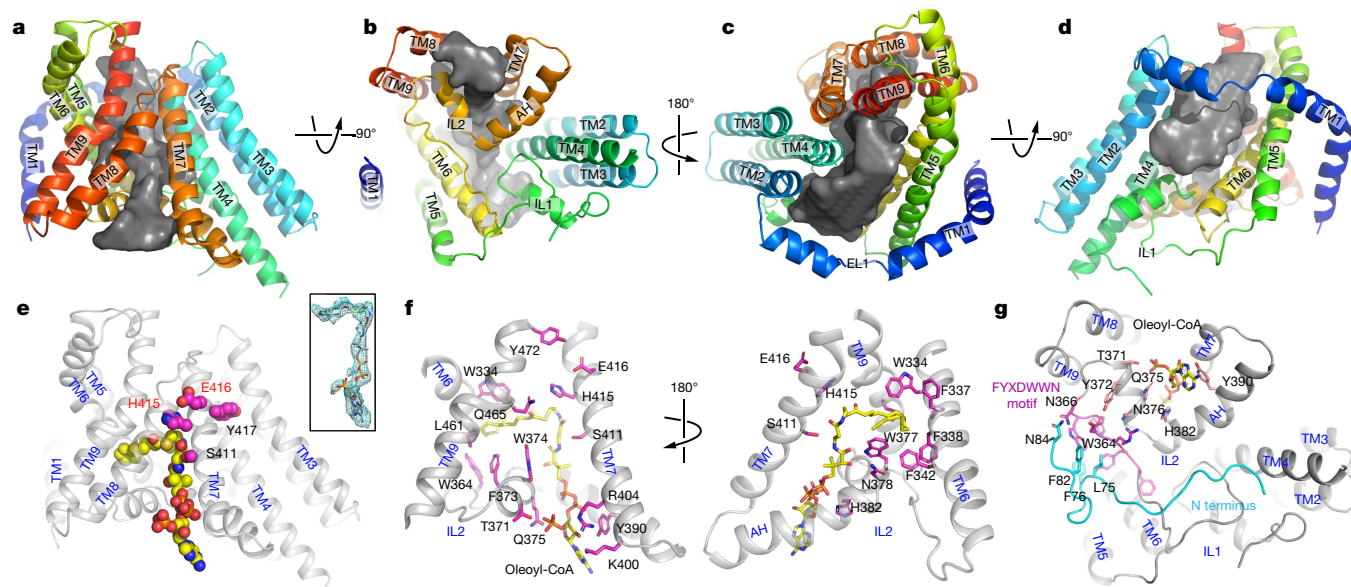


Fig. 2 | The reaction chamber and oleoyl-CoA-binding site. **a–d**, The reaction chamber (grey surface) is shown in four orientations with the surrounding helices in cartoon representation. **e**, An oleoyl-CoA molecule is shown as spheres with carbon atoms coloured in yellow. The side chains of the conserved active site residues (SXXHEY) are shown as magenta spheres. Inset, oleoyl-CoA

is shown as sticks and its density as green mesh. **f**, Residues at the oleoyl-CoA-binding site are shown as sticks with carbon atoms coloured in magenta. AH, amphipathic helix. **g**, Interaction between the FYXDWWN motif (magenta) and the N terminus of the neighbouring protomer (cyan).

N terminus before TM1 led to loss of the enzymatic activity^{18–20}. The current structure shows that although the N terminus is not structured, it interacts with the highly conserved elements in the MBOAT fold. We next examined the functional us before TM1 led to loss of the enzymatic activi implications of the domain-swapped N terminus of DGAT1. We progressively shortened the N terminus by constructing DGAT1 mutants with deletions of residues 2–65 (Δ N65), 2–70 (Δ N70), 2–75 (Δ N75), 2–80 (Δ N80) and 2–84 (Δ N84), and were able to purify all of these as stable dimers (Extended Data Fig. 5f). We found that the enzymatic activity is progressively lower as more N-terminal residues are deleted, and the dimer with the longest N-terminal deletion (Δ N84) has no activity (Fig. 1j, Extended Data Table 2). These results indiate that the N terminus could regulate enzymatic function by its interaction with the MBOAT fold.

The reaction chamber

The MBOAT fold of DGAT1 (TM2–TM9, IL1 and IL2) carves out a large hollow chamber in the hydrophobic core of the membrane (Fig. 2a–d). His415, which is almost universally conserved in the MBOAT family of enzymes, is found inside the reaction chamber and on TM7. TM2–TM9 segregate into three groups that form three sidewalls of the chamber. TM2, TM3 and TM4 pack into a bundle that forms the first sidewall; TM5 and TM6 are both very long (almost 40 amino acids each) and the two helices coil into a unit that tilts roughly 56° to the membrane norm to form the second sidewall; and TM7, TM8 and TM9 form a panel and the third sidewall (Fig. 2a–d). The cytosolic ends of TM7 and TM8 are around 19 Å apart, creating a cytosolic entrance to the reaction chamber (Fig. 2a, b). IL1 and IL2 form the floor of the chamber at the cytosolic side. IL1 (residues 222 to 261) is composed of a helix flanked by two long strands, whereas IL2 (residues 352 to 396) has a long amphipathic helix (residues 380 to 394) preceded by a short helix and a loop.

Acyl-CoA-binding site

The structure of human DGAT1 was solved in the presence of 2 mM oleoyl-CoA. A large non-protein density is found at the cytosolic side of

the reaction chamber close to IL2 and it extends deep into the reaction chamber (Extended Data Fig. 4b). An oleoyl-CoA molecule is modelled into this density, with the adenosine 3',5'-diphosphate of the CoA moiety at the cytosolic entrance, the 4-phosphate panthothenic acid, β -alanine and β -mercapto-ethylamine extending progressively into the reaction chamber and the acyl chain residing in a hydrophobic pocket inside the reaction chamber (Fig. 2e, f, Extended Data Fig. 6a–h). The activated thioester is located in the vicinity of His415, poised for an attack from the hydroxyl group of DAG. The position of the thioester could be stabilized by interactions between the carbonyl oxygen of the fatty acid and the side chain of Gln465 on TM9 (Fig. 2f, Extended Data Fig. 6h).

IL2 has a crucial role in acyl-CoA binding. The V-shaped helix-turn-helix motif of IL2 forms the binding site for the adenosine 3',5'-diphosphate moiety of acyl-CoA (Fig. 2g). The loop preceding the helix-turn-helix motif contains a highly conserved FYXDWWN sequence that is found in both DGAT1 and the related ACAT, and mutational studies suggest that these residues may coordinate acyl-CoA^{2,21} (Extended Data Fig. 7a, c). However, only Trp364, the first tryptophan in the FYXDWWN sequence, forms part of the hydrophobic pocket for the acyl chain and the rest of the sequence does not have direct contact with the acyl-CoA. FYXDWWN packs tightly against the helix-turn-helix motif of IL2 and also interacts extensively with the N terminus from the neighbouring protomer (Fig. 2g). We speculate that mutations to this sequence and deletion of the N terminus could affect the enzymatic activity by perturbing these interactions.

To assess the functional effect of residues at the active site and those that line the acyl-CoA-binding site, we introduced point mutations and measured the enzymatic activities of these different DGAT1 mutants. Point mutations to residues that line the entrance of the acyl-CoA-binding site (Thr371, Tyr390, Lys400 and Arg404) reduce the enzymatic activity by 30–70%, whereas point mutations to residues of the rest of the binding pocket (Trp377, Asn378, His382 and Ser411) have a larger effect, resulting in a loss of enzymatic activity of more than 80%. Mutations to residues of the active site (His415 and Glu416) abolish enzymatic activity (Extended Data Fig. 6i, j).

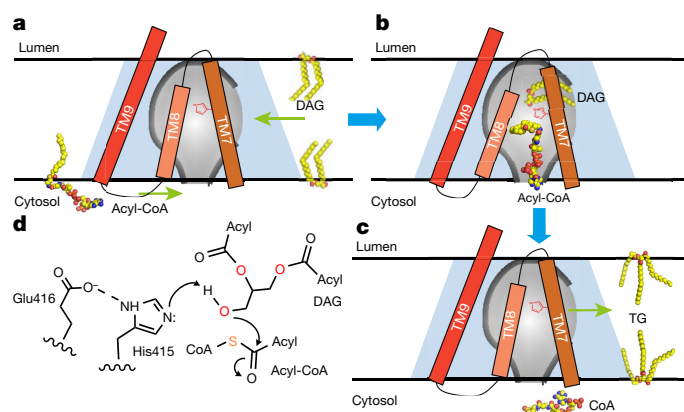


Fig. 3 | Proposed catalytic mechanism of human DGAT1. **a, b**, ADGAT1 monomer is shown as a trapezoid in light blue and the reaction chamber is shown in the shape of an inverted flask coloured in grey. TM7–TM9, acyl-CoA and DAG are shown schematically. The catalytic His415 residue is marked in red on TM7. The CoA moiety of an acyl-CoA molecule binds to DGAT1 at the cytosolic entrance of the tunnel and the hydrophobic acyl chain slides into the reaction chamber through a slit between TM7 and TM8. The glycerol backbone of DAG enters the chamber from the large side entrance with the two acyl chains partially hosted in the hydrophobic core of the membrane. **c**, After the reaction, the product triacylglyceride (TG) could diffuse into either leaflet of the membrane. **d**, Proposed catalytic mechanism. E416 and H415 activate the 3-hydroxyl group on DAG for a nucleophilic attack on the thioester of the acyl-CoA molecule.

Gateway for DAG and triacylglyceride

The reaction chamber has a large lateral opening to the hydrophobic core of the membrane; this opening is framed by TM4 on one side and TM6 on the other side, and by part of IL1 (residues 234–245) on the cytosolic side (Fig. 2c, d, Extended Data Fig. 8a–c). Residues that line the two sides of the entrance are mostly hydrophobic (Extended Data Fig. 8d). A tubular density is found near the entrance and extends deep into the reaction chamber (Extended Data Figs. 4b, 8a). Residues surrounding the tubular density are mostly hydrophobic, indicating that it is probably a long aliphatic acyl chain, although the density is not large enough to accommodate an extended DAG. We speculate that the lateral opening would allow the entrance of DAG and exit of triacylglyceride—both of which can be accommodated by the hydrophobic core of the membrane. Consistent with this hypothesis, mutating Leu346 to a bulkier tryptophan side chain produces an enzyme that has less than 10% the activity of the wild type (Extended Data Fig. 6i, j).

Discussion

The structure of human DGAT1 defines a conserved MBOAT structural fold, which forms a reaction chamber in the ER membrane to shield the acyl transfer reaction from the hydrophobic core of the membrane. The reaction chamber has a tunnel to the cytosolic side and its entrance recognizes the hydrophilic coenzyme A motif of an acyl-CoA molecule. The tunnel has a slit between TM7 and TM8 that could allow the entry of the acyl chain of an acyl-CoA molecule into the chamber, reaching the hydrophobic pocket inside the chamber (Fig. 3a, b). The reaction chamber also has a large opening to the hydrophobic core of the membrane, which could allow entry of a DAG molecule. We propose that when the

glycerol backbone of a DAG molecule approaches the catalytic centre at His415, the two hydrophobic aliphatic acyl chains of DAG could remain partially outside of the protein, accommodated in the hydrophobic core of the membrane (Fig. 3b). The conserved His415 would facilitate the acyl transfer reaction by activating the free hydroxyl group on DAG, and Glu416 could enhance the activation. The activated hydroxyl oxygen then attacks the thioester on the fatty acyl-CoA to form a new ester bond (Fig. 3d). The product, triacylglyceride, could return into the hydrophobic core of the membrane while coenzyme A dissociates into the cytosol (Fig. 3c).

Online content

Any methods, additional references, Nature Research reporting summaries, source data, extended data, supplementary information, acknowledgements, peer review information; details of author contributions and competing interests; and statements of data and code availability are available at <https://doi.org/10.1038/s41586-020-2280-2>.

- Cases, S. et al. Identification of a gene encoding an acyl CoA:diacylglycerol acyltransferase, a key enzyme in triacylglycerol synthesis. *Proc. Natl Acad. Sci. USA* **95**, 13018–13023 (1998).
- Das, A., Davis, M. A. & Rudel, L. L. Identification of putative active site residues of ACAT enzymes. *J. Lipid Res.* **49**, 1770–1781 (2008).
- Nusse, R. Wnt signaling in disease and in development. *Cell Res.* **15**, 28–32 (2005).
- Yen, C.-L. E., Nelson, D. W. & Yen, M.-I. Intestinal triacylglycerol synthesis in fat absorption and systemic energy metabolism. *J. Lipid Res.* **56**, 489–501 (2015).
- Yen, C.-L. E., Stone, S. J., Koliwad, S., Harris, C. & Farese, R. V. Jr. Thematic review series: glycerolipids. DGAT enzymes and triacylglycerol biosynthesis. *J. Lipid Res.* **49**, 2283–2301 (2008).
- Bhatt-Wessel, B., Jordan, T. W., Miller, J. H. & Peng, L. Role of DGAT enzymes in triacylglycerol metabolism. *Arch. Biochem. Biophys.* **655**, 1–11 (2018).
- Kühn, C. et al. Evidence for multiple alleles at the DGAT1 locus better explains a quantitative trait locus with major effect on milk fat content in cattle. *Genetics* **167**, 1873–1881 (2004).
- Smith, S. J. et al. Obesity resistance and multiple mechanisms of triglyceride synthesis in mice lacking Dgat. *Nat. Genet.* **25**, 87–90 (2000).
- Chen, H. C. et al. Increased insulin and leptin sensitivity in mice lacking acyl CoA:diacylglycerol acyltransferase 1. *J. Clin. Invest.* **109**, 1049–1055 (2002).
- Zhao, G. et al. Validation of diacyl glycerol acyltransferase 1 as a novel target for the treatment of obesity and dyslipidemia using a potent and selective small molecule inhibitor. *J. Med. Chem.* **51**, 380–383 (2008).
- McFie, P. J., Stone, S. L., Banman, S. L. & Stone, S. J. Topological orientation of acyl-CoA:diacylglycerol acyltransferase-1 (DGAT1) and identification of a putative active site histidine and the role of the N terminus in dimer/tetramer formation. *J. Biol. Chem.* **285**, 37377–37387 (2010).
- Yang, J., Brown, M. S., Liang, G., Grishin, N. V. & Goldstein, J. L. Identification of the acyltransferase that octanoylates ghrelin, an appetite-stimulating peptide hormone. *Cell* **132**, 387–396 (2008).
- Ma, D. et al. Crystal structure of a membrane-bound O-acyltransferase. *Nature* **562**, 286–290 (2018).
- Cheng, D. et al. Human acyl-CoA:diacylglycerol acyltransferase is a tetrameric protein. *Biochem. J.* **359**, 707–714 (2001).
- Zhang, J. et al. Monoacylglycerol acyltransferase-2 is a tetrameric enzyme that selectively heterodimerizes with diacylglycerol acyltransferase-1. *J. Biol. Chem.* **289**, 10909–10918 (2014).
- Cao, J. et al. Targeting acyl-CoA:diacylglycerol acyltransferase 1 (DGAT1) with small molecule inhibitors for the treatment of metabolic diseases. *J. Biol. Chem.* **286**, 41838–41851 (2011).
- von Heijne, G. & Gavel, Y. Topogenic signals in integral membrane proteins. *Eur. J. Biochem.* **174**, 671–678 (1988).
- Caldo, K. M. P. et al. Diacylglycerol acyltransferase 1 is regulated by its N-terminal domain in response to allosteric effectors. *Plant Physiol.* **175**, 667–680 (2017).
- Weslake, R. J. et al. Acyl-CoA-binding and self-associating properties of a recombinant 13.3 kDa N-terminal fragment of diacylglycerol acyltransferase-1 from oilseed rape. *BMC Biochem.* **7**, 24 (2006).
- Panigrahi, R. et al. Intrinsic disorder in the regulatory N-terminal domain of diacylglycerol acyltransferase 1 from *Brassica napus*. *Sci. Rep.* **8**, 16665 (2018).
- Guo, Z., Cromley, D., Billheimer, J. T. & Sturley, S. L. Identification of potential substrate-binding sites in yeast and human acyl-CoA sterol acyltransferases by mutagenesis of conserved sequences. *J. Lipid Res.* **42**, 1282–1291 (2001).

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

© The Author(s), under exclusive licence to Springer Nature Limited 2020

Methods

Data reporting

No statistical methods were used to predetermine sample size. The experiments were not randomized and the investigators were not blinded to allocation during experiments and outcome assessment.

Cloning, expression and purification of human DGAT1

The human *DGAT1* gene (accession number NP_036211) was codon-optimized and cloned into a modified pFastBac Dual vector²² for production of baculovirus by the Bac-to-Bac method (Thermo Fisher Scientific). High Five cells (Thermo Fisher Scientific) at a density of around 3×10^6 cells ml^{-1} were infected with baculovirus and grown at 27 °C for 48–60 h before collection. Cell membranes were isolated following a previous protocol²² and flash-frozen in liquid nitrogen.

Isolated cell membranes were thawed and homogenized in 20 mM HEPES, pH 7.5, 150 mM NaCl and 2 mM β -mercaptoethanol, and then solubilized with 1% (w/v) LMNG (Anatrace) at 4 °C for 2 h. After centrifugation (55,000g, 45 min, 4 °C), DGAT1 was purified from the supernatant using a cobalt-based affinity resin (Talon, Clontech) and the His₆-tag was cleaved by incubation with tobacco etch virus (TEV) protease for 1 h at room temperature. Oleoyl-CoA (20 μM) was added to reduce aggregation, and DGAT1 was then concentrated to around 5 mg ml^{-1} (Amicon 100 kDa cut-off, Millipore) and loaded onto a size-exclusion column (SRT-3C SEC-300, Sepax Technologies) equilibrated with 20 mM HEPES, pH 7.5, 150 mM NaCl and 0.02% GDN (Anatrace). Purified DGAT1 was mixed with 2 mM oleoyl-CoA and concentrated to around 20 mg ml^{-1} for cryo-EM grid preparation.

When LMNG is used in the extraction step, most of the DGAT1 is homodimer and only a small fraction is homotetramer. To obtain tetrameric DGAT1, 1% GDN was used for extraction and 0.02% GDN for purification. The dimeric DGAT1 produces substantially better cryo-EM grids and was given priority for structure determination.

DGAT1 mutants were generated using the QuikChange method and the entire cDNA was sequenced to verify the mutation. Mutants were expressed and purified following the same protocol as wild type.

Cryo-EM sample preparation and data collection

Cryo grids were prepared using the Thermo Fisher Vitrobot Mark IV. Quantifoil R1.2/1.3 Cu grids were glow-discharged in air for 40 s at medium level using the Plasma Cleaner (Harrick Plasma, PDC-32G-2). Concentrated DGAT1 (3.5 μl) was applied to each glow-discharged grid. After blotting with filter paper (Ted Pella, Prod. 47000-100) for 3.5 s, the grids were plunged into liquid ethane cooled with liquid nitrogen. Movie stacks were collected using SerialEM²³ on a Titan Krios at 300 kV with a Quantum energy filter (Gatan) and a Cs corrector (Thermo Fisher Scientific), at a nominal magnification of $\times 105,000$ and with defocus values of $-2.0 \mu\text{m}$ to $-1.2 \mu\text{m}$. A K2 Summit direct electron detector (Gatan) was paired with the microscope. Each stack was collected in the super-resolution mode with an exposing time of 0.175 s per frame for a total of 32 frames. The dose was about 50 e^- per \AA^2 for each stack. The stacks were motion-corrected with MotionCor2²⁴ and binned (2×2) so that the pixel size was 1.114 \AA . Dose weighting²⁵ was performed during motion correction, and the defocus values were estimated with Gctf²⁶.

Cryo-EM data processing

A total of 2,749,110 particles were automatically picked (RELION 2.1^{27–29}) from 3,510 images and imported into cryoSPARC³⁰. Out of 200 two-dimensional (2D) classes, 101 (containing 1,000,063) particles were selected for ab initio three-dimensional (3D) reconstruction, which produced one good class with recognizable structural features and three bad classes that do not have structural features (Extended Data Fig. 3). Although human DGAT1 can form both dimers and tetramers, as only the dimer fraction was used in grid preparation

we found no tetramer during 2D classification. Both the good and bad classes were used as references in the heterogeneous refinement (cryoSPARC) and yielded a good class at 4.1 \AA from 408,945 particles. After handedness correction, non-uniform refinement (cryoSPARC) was performed with C2 symmetry and an adaptive solvent mask, which yielded a map with an overall resolution of 3.1 \AA (Extended Data Fig. 3b). Further heterogeneous refinement yielded a class with 275,945 particles and after non-uniform refinement, a map was yielded that had similar resolution but improved density of TM2, TM3, TM8 and lipids (Extended Data Fig. 3c). Resolutions were estimated using the gold-standard Fourier shell correlation with a 0.143 cut-off³¹ and high-resolution noise substitution³². Local resolution was estimated using ResMap³³.

Model building and refinement

Structure models were built de novo into the density map starting with poly-alanine, and side chains were then added onto the model based on the map. Model building was conducted in Coot³⁴. Structure refinements were carried out in PHENIX in real space with secondary structure and geometry restraints³⁵. The EMRinger Score was calculated as described³⁶.

DGAT1 activity assay

DGAT1 activity was measured using a fluorescence-based coupled-enzyme assay³⁷ in a quartz cuvette at 37 °C (Extended Data Fig. 2d). The reaction was monitored in a FluoroMax-4 spectrofluorometer (HORIBA) with 340-nm excitation and 465-nm emission at 15-s intervals. All assays were done in a buffer with 20 mM HEPES, pH 7.5, 150 mM NaCl, 2 mM β -mercaptoethanol, 0.5 mM DDM and 1% TritonX-100. Final concentrations of NAD⁺, thiamine pyrophosphate and α -ketoglutarate were 0.25 mM, 0.2 mM and 2 mM, respectively. α -ketoglutarate dehydrogenase (αKDH) was prepared from bovine heart purchased from a meat market, following a published protocol³⁸. An appropriate amount of αKDH was used to ensure that the DGAT1 reaction is the rate-limiting step. When oleoyl-CoA concentrations were varied, the DAG concentration was fixed at 200 μM . When DAG concentrations were varied, the oleoyl-CoA concentration was fixed at 100 μM . All reactions were initiated with the addition of oleoyl-CoA. The initial rate versus different concentrations of oleoyl-CoA can be fit with a Michaelis–Menten equation. The initial rates in various concentrations of DAG were not well fit with the traditional Michaelis–Menten equation, but could be fit with an allosteric sigmoidal equation: $Y = V_{\text{max}} \times X^h / (K_m + X^h)$, in which X is the DAG concentration and h is the Hill coefficient.

When assaying the activity of DGAT1 dimer or tetramer, the protein concentration was kept at 2.4 $\mu\text{g ml}^{-1}$ (around 40 nM). When measuring DGAT1 in the cell membrane, crude membrane containing DGAT1 was used and the amount of DGAT1 in the membrane was estimated on the basis of the yield of DGAT1 from the same batch of cells. We did not observe substrate inhibition up to 200 μM of oleoyl-CoA.

Detection of triacylglycerides by thin-layer chromatography

To validate the functional assay described in the previous section, we confirmed triacylglyceride production directly. After initiating an enzymatic reaction, 100 μl of the sample was taken at each indicated time point and extracted with 400 μl chloroform. The organic phase containing triacylglyceride was dried under argon and then resuspended in 40 μl chloroform out of which 4 μl was spotted onto a KC18 reversed-phase thin-layer chromatography plate (Whatman Chemical Separation). The mobile phase is 100:1 (chloroform: acetic acid, v/v) and triacylglyceride was visualized in an I₂ chamber.

Reporting summary

Further information on research design is available in the Nature Research Reporting Summary linked to this paper.

Data availability

The atomic coordinates of human DGAT1 have been deposited in the Protein Data Bank (<http://www.rcsb.org>) under the accession code 6VP0. The corresponding electron microscopy maps have been deposited in the Electron Microscopy Data Bank (<https://www.ebi.ac.uk/pdbe/emdb/>) under the accession code EMD-21302

22. Bai, Y. et al. X-ray structure of a mammalian stearyl-CoA desaturase. *Nature* **524**, 252–256 (2015).
23. Mastronarde, D. N. Automated electron microscope tomography using robust prediction of specimen movements. *J. Struct. Biol.* **152**, 36–51 (2005).
24. Zheng, S. Q. et al. MotionCor2: anisotropic correction of beam-induced motion for improved cryo-electron microscopy. *Nat. Methods* **14**, 331–332 (2017).
25. Grant, T. & Grigorieff, N. Measuring the optimal exposure for single particle cryo-EM using a 2.6 Å reconstruction of rotavirus VP6. *eLife* **4**, e06980 (2015).
26. Zhang, K. Gctf: real-time CTF determination and correction. *J. Struct. Biol.* **193**, 1–12 (2016).
27. Scheres, S. H. Semi-automated selection of cryo-EM particles in RELION-1.3. *J. Struct. Biol.* **189**, 114–122 (2015).
28. Scheres, S. H. RELION: implementation of a Bayesian approach to cryo-EM structure determination. *J. Struct. Biol.* **180**, 519–530 (2012).
29. Kimanius, D., Forsberg, B. O., Scheres, S. H. & Lindahl, E. Accelerated cryo-EM structure determination with parallelisation using GPUs in RELION-2. *eLife* **5**, e18722 (2016).
30. Punjani, A., Rubinstein, J. L., Fleet, D. J. & Brubaker, M. A. cryoSPARC: algorithms for rapid unsupervised cryo-EM structure determination. *Nat. Methods* **14**, 290–296 (2017).
31. Rosenthal, P. B. & Henderson, R. Optimal determination of particle orientation, absolute hand, and contrast loss in single-particle electron cryomicroscopy. *J. Mol. Biol.* **333**, 721–745 (2003).
32. Chen, S. et al. High-resolution noise substitution to measure overfitting and validate resolution in 3D structure determination by single particle electron cryomicroscopy. *Ultramicroscopy* **135**, 24–35 (2013).
33. Kucukelbir, A., Sigworth, F. J. & Tagare, H. D. Quantifying the local resolution of cryo-EM density maps. *Nat. Methods* **11**, 63–65 (2014).
34. Emsley, P., Lohkamp, B., Scott, W. G. & Cowtan, K. Features and development of Coot. *Acta Crystallogr. D* **66**, 486–501 (2010).
35. Adams, P. D. et al. PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr. D* **66**, 213–221 (2010).
36. Barad, B. A. et al. EMRinger: side chain-directed model and map validation for 3D cryo-electron microscopy. *Nat. Methods* **12**, 943–946 (2015).
37. Rana, M. S. et al. Fatty acyl recognition and transfer by an integral membrane S-acyltransferase. *Science* **359**, eaao6326 (2018).
38. Stanley, C. J. & Perham, R. N. Purification of 2-oxo acid dehydrogenase multienzyme complexes from ox heart by a new method. *Biochem. J.* **191**, 147–154 (1980).

39. Ren, Z. et al. Structure of an EIIC sugar transporter trapped in an inward-facing conformation. *Proc. Natl Acad. Sci. USA* **115**, 5962–5967 (2018).
40. Baker, N. A., Sept, D., Joseph, S., Holst, M. J. & McCammon, J. A. Electrostatics of nanosystems: application to microtubules and the ribosome. *Proc. Natl Acad. Sci. USA* **98**, 10037–10041 (2001).
41. Wallace, A. C., Laskowski, R. A. & Thornton, J. M. LIGPLOT: a program to generate schematic diagrams of protein-ligand interactions. *Protein Eng. Des. Sel.* **8**, 127–134 (1995).
42. Laskowski, R. A. & Swindells, M. B. LigPlot+: multiple ligand-protein interaction diagrams for drug discovery. *J. Chem. Inf. Model.* **51**, 2778–2786 (2011).
43. Sievers, F. et al. Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. *Mol. Syst. Biol.* **7**, 539 (2011).
44. Robert, X. & Gouet, P. Deciphering key features in protein structures with the new ENDscript server. *Nucleic Acids Res.* **42**, W320–W324 (2014).
45. Landau, M. et al. ConSurf 2005: the projection of evolutionary conservation scores of residues on protein structures. *Nucleic Acids Res.* **33**, W299–W302 (2005).

Acknowledgements This work was supported by grants from the NIH (DK122784, HL086392 and GM098878 to M.Z.), the Cancer Prevention and Research Institute of Texas (R1223 to M.Z.), the Robert Welch Foundation (Q1279 to B.V.V.P.), the Ara Parseghian Medical Research Foundation (to N.Y. and Y.H.) and the New Jersey Council for Cancer Research (to H.Q.). N.Y. is supported by the Shirley M. Tilghman endowed professorship from Princeton University. We thank P. Shao for technical support during the acquisition of electron microscope images, and A. Banerjee for advice on the enzymatic assay and purification of αKDH. We acknowledge the use of the Imaging and Analysis Center at Princeton University, which is partially supported by the Princeton Center for Complex Materials, and the National Science Foundation (NSF)-MRSEC program (DMR-1420541).

Author contributions M.Z., L.W., Y.N. and Z.R. conceived the project and designed the experiments. L.W., Y.N. and H.Z. expressed and purified human DGAT1. L.W. and Y.N. measured functions of DGAT1. H.Q., Y.H., L.W., Z.R. and H.Z. prepared cryo-EM grids and collected and analysed cryo-EM data. L.W. and Z.R. built and refined the human DGAT1 structure. L.H. and B.V.V.P. advised on model building and refinement. A.L. examined the lipid content of DGAT1. L.W., Z.R., N.Y. and M.Z. analysed data. L.W., Z.R. and M.Z. wrote the initial draft and all authors participated in revising the manuscript.

Competing interests The authors declare no competing interests.

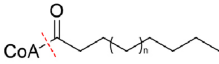
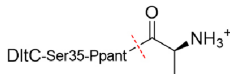
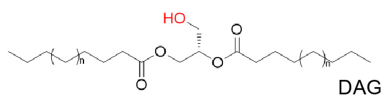
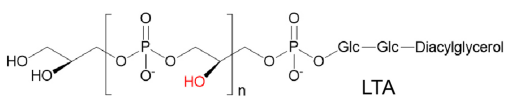
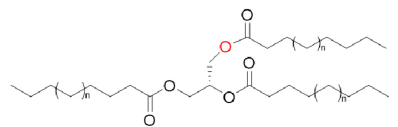
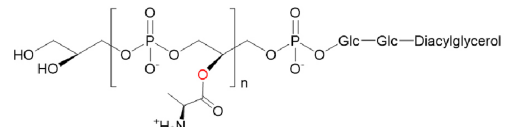
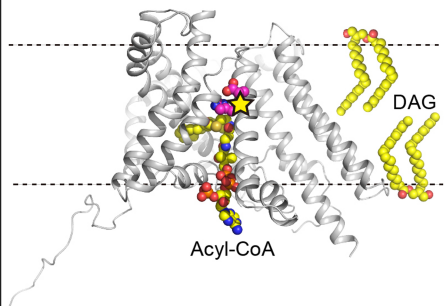
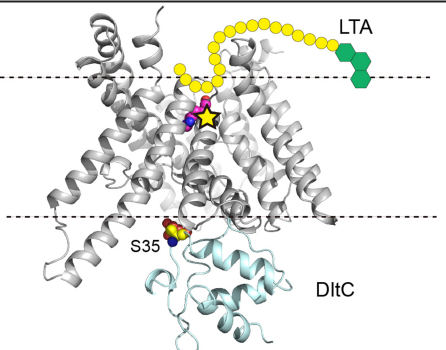
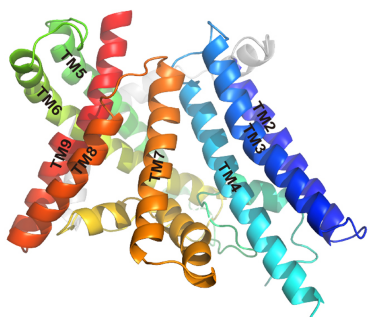
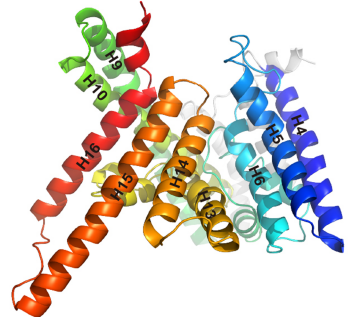
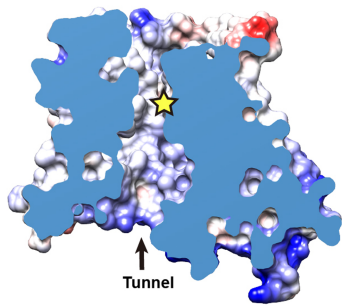
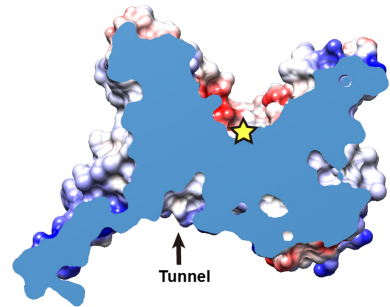
Additional information

Supplementary information is available for this paper at <https://doi.org/10.1038/s41586-020-2280-2>.

Correspondence and requests for materials should be addressed to N.Y. or M.Z.

Peer review information *Nature* thanks David Drew, Savvas N. Savvides and the other, anonymous, reviewer(s) for their contribution to the peer review of this work.

Reprints and permissions information is available at <http://www.nature.com/reprints>.

MBOAT	DGAT	DltB
Acyl-donor	 Acyl-CoA	 DltC-Ser35-Pant
Acyl-acceptor	 DAG	 LTA
Product		
Substrates distribution	 Acyl-CoA	 S35
MBOAT fold		
Tunnel	 Tunnel	 Tunnel

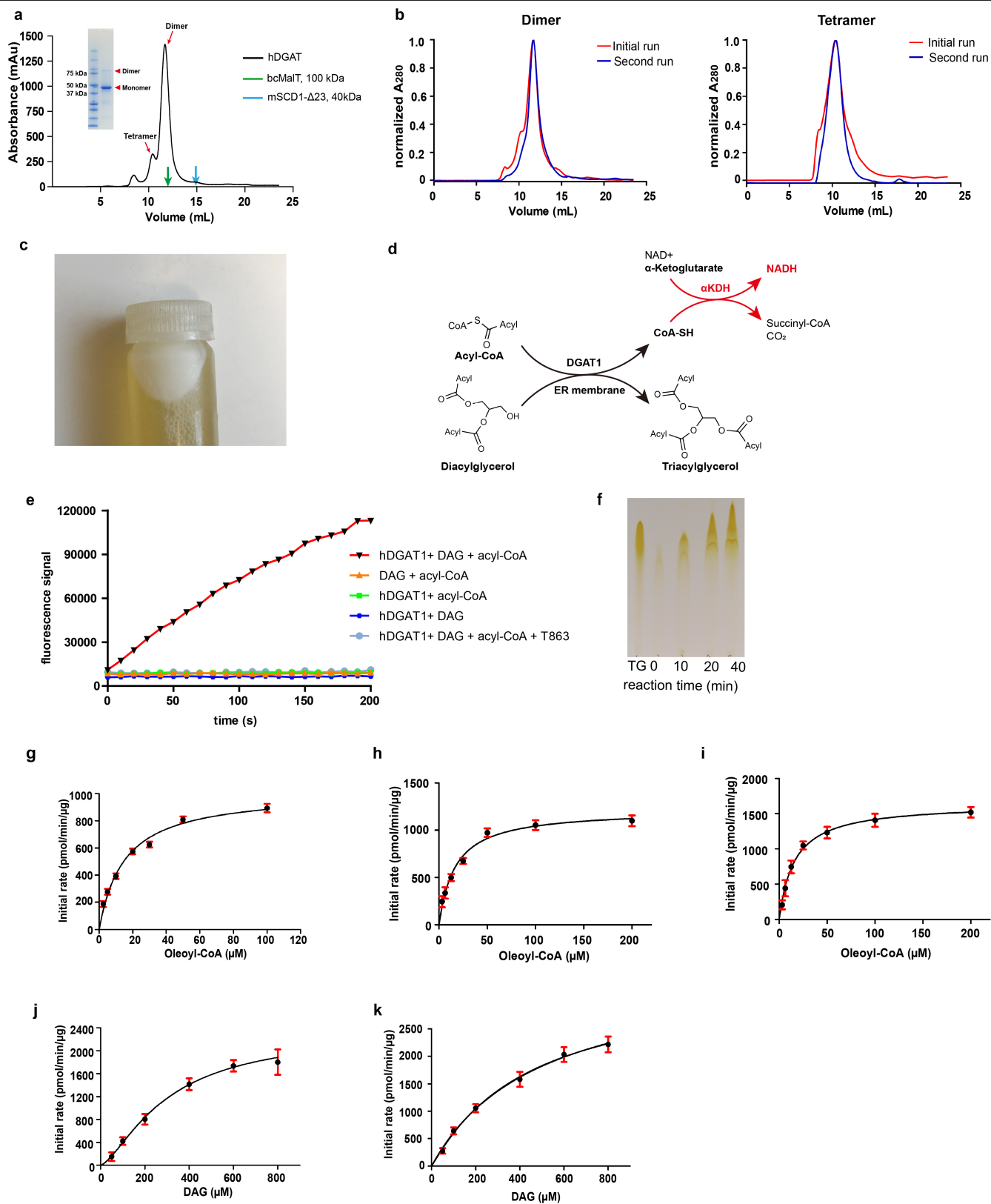
Extended Data Fig. 1 | See next page for caption.

Article

Extended Data Fig. 1 | Side-by-side comparison of human DGAT1 and DltB.

Both human DGAT1 and DltB have an acyl donor and an acyl acceptor. In the acyl-donor row, the red dashed lines indicate the bonds that are broken during acyl-transfer reactions. In the acyl-acceptor row, the hydroxyl groups are highlighted in red. In the substrates distribution row, DGAT1 and the DltB–DltC complex are shown as cartoon and the membrane as dashed lines. The position of the catalytic histidine in each protein is marked with a yellow star. In human DGAT1, the acyl-CoA comes from the intracellular side whereas the DAG comes from the hydrophobic core of the membrane. In DltB, the 4'-phosphopantetheine-DltC is intracellular whereas the lipoteichoic acid (LTA) is extracellular. In the MBOAT fold row, the MBOAT folds of DGAT1 and

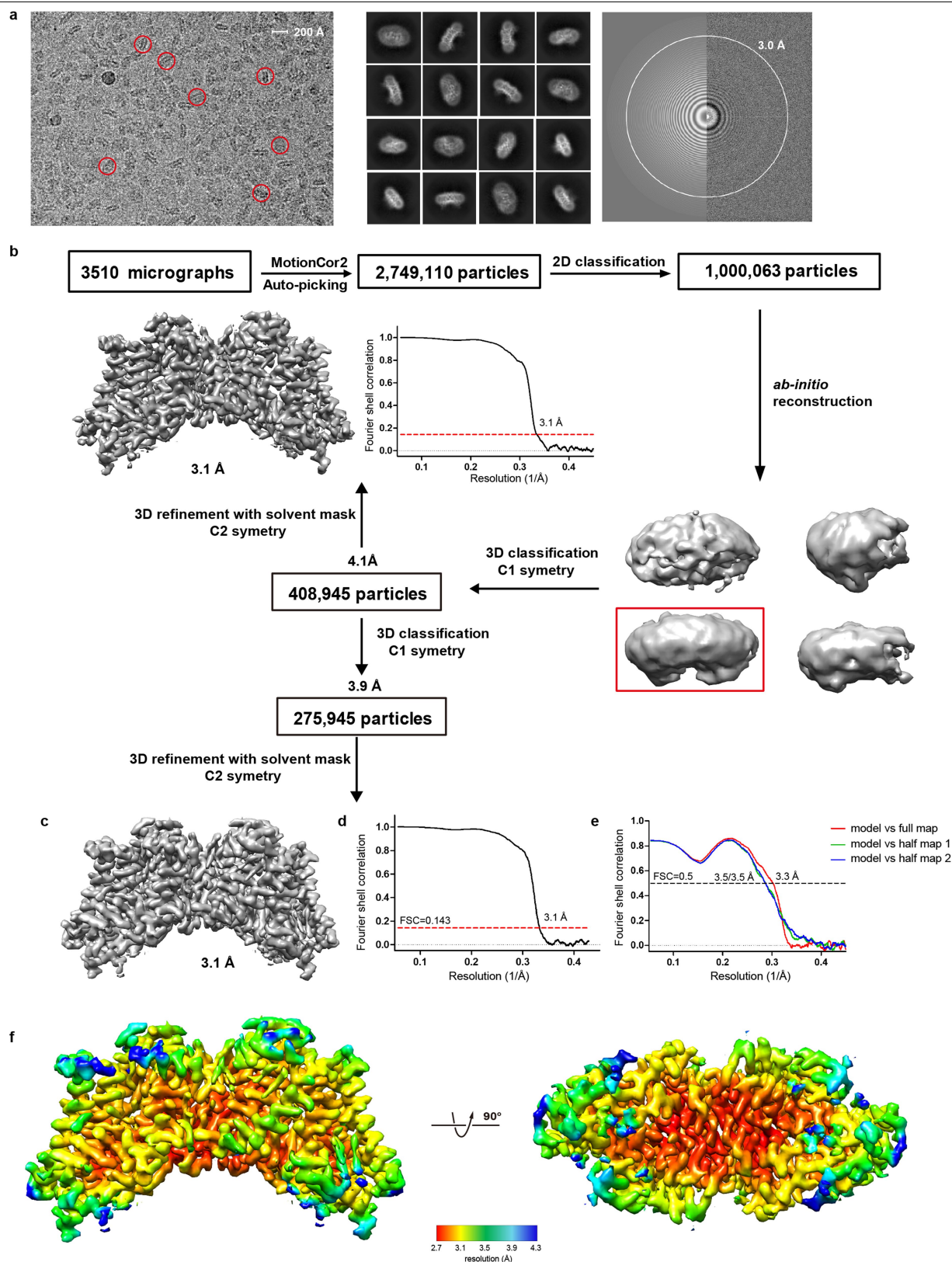
DltB are shown in cartoon representation and viewed from the same orientation. Equivalent helices have the same colour. The tunnel row shows the cut-away surface illustrations of DGAT1 and DltB, showing their cytosolic tunnels. The position of the conserved histidine residue in each protein is marked with a yellow star. In DltB, the intracellular loops are placed more towards the centre of the membrane and as a result, the MBOAT fold in DltB does not carve out a reaction chamber in the membrane. Overall, DltB is shaped similarly to an hourglass that allows the two substrates to approach the reaction centre from either side of the membrane, and the transfer of an acyl group across the membrane. These observations highlight the versatility of the MBOAT fold.



Extended Data Fig. 2 | See next page for caption.

Extended Data Fig. 2 | Purification and functional characterization of human DGAT1. **a**, Size-exclusion chromatograph profile of human DGAT1 extracted with LMNG. Elution volumes of membrane proteins of known molecular weight—bcMalT (about 100 kDa, green)³⁹ and mouse (m)SCD1 (about 40 kDa, blue)²²—are marked with arrows. Inset, SDS–PAGE of the purified DGAT1. DGAT1 has a main peak with an elution volume of around 11.7 ml that corresponds to a dimer and a minor peak at around 10.4 ml that corresponds to a tetramer. **b**, Size-exclusion chromatography profiles of DGAT1 extracted with LMNG (left) or GDN (right). The main peak from the first run (red trace) was collected and reinjected onto the same column after 1 h. In both **a** and **b**, the detergent in the mobile phase is GDN. **c**, A white layer of fat appeared after membrane solubilization and centrifugation, indicating that the heterologously expressed DGAT1 is active in cells. **d**, The DGAT1 reaction is coupled to that of α KDH to monitor production of coenzyme A (CoA-SH) in real time. **e**, Fluorescence of NADH plotted versus time. Rapid production of

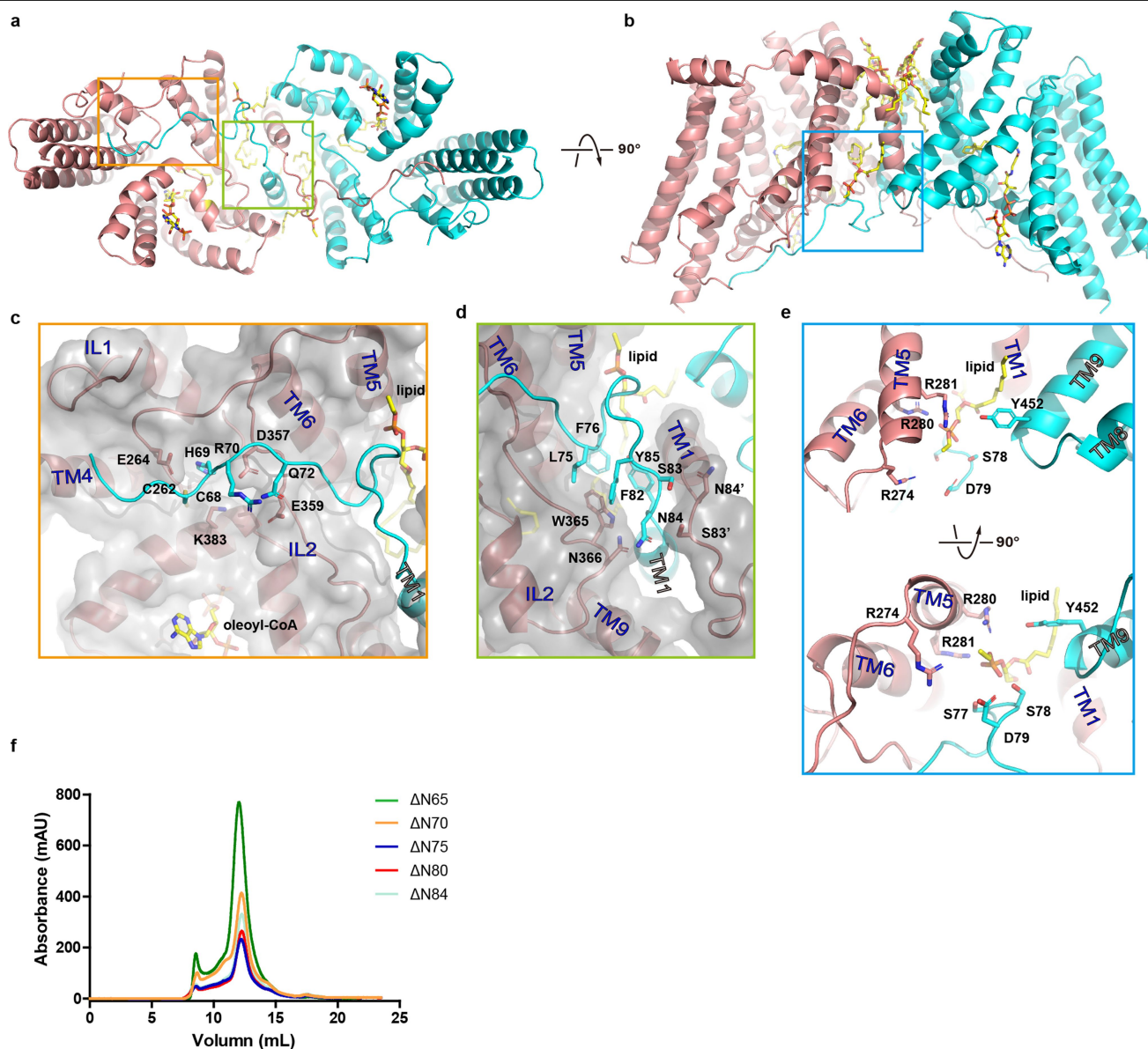
coenzyme A occurs in the presence of oleoyl-CoA, 1,2-dioleoyl-sn-glycerol (1,2-DAG) and the purified dimeric DGAT1. By contrast, production of coenzyme A was not observed when either 1,2-DAG or DGAT1 was omitted from the reaction mixture, indicating that hydrolysis of oleoyl-CoA is tightly coupled to the enzymatic reaction. In addition, coenzyme A production was almost completely suppressed in the presence of 5 μ M T863, a known DGAT1 inhibitor¹⁶. **f**, Production of triacylglyceride over time, detected by thin-layer chromatography. The first lane from the left is a triacylglyceride standard. **g–i**, Initial rate of reaction versus oleoyl-CoA concentration measured using the purified dimeric DGAT1 (**g**), tetrameric DGAT1 (**h**) or DGAT1 in cell membrane (**i**). **j, k**, Initial rate of reaction versus DAG concentration measured using the dimeric (**j**) or tetrameric (**k**) DGAT1. Data are mean \pm s.e.m. derived from three independent repeats. Experiments were repeated independently 10 times with similar results (**a, c**) or 3 times with similar results (**b, e, f**).



Extended Data Fig. 3 | Cryo-EM data and processing. **a**, A representative micrograph of DGAT1 (left), its Fourier transform (right) and representative 2D class averages (middle). Representative particles are highlighted with red circles. **b, c**, A flow chart for data processing and the final maps of

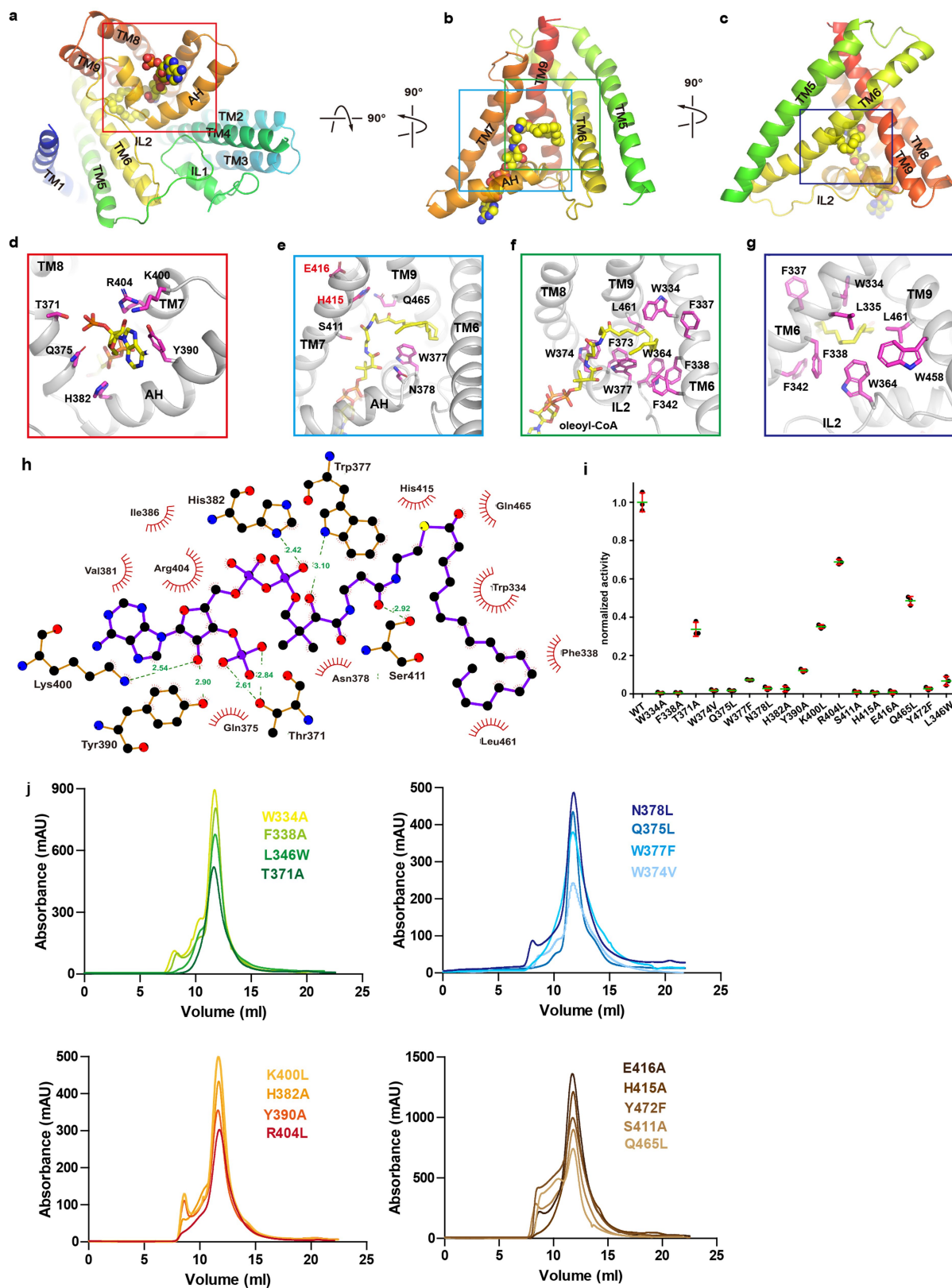
DGAT1 (Methods). **d**, The gold-standard Fourier shell correlation (FSC) curve for the final map shown in **c**. **e**, Fourier shell correlation curves of the atomic model of DGAT1 versus the full map and individual half maps. **f**, Local-resolution map of DGAT1 shown in two orientations.

(green mesh) is shown at the same contour level as its neighbouring helix, TM4 (red mesh). **c**, Detailed view of each detergent or lipid molecule and its density. **d**, Electrostatic surface representations of the DGAT1 dimer in three orientations. The electrostatic potential was calculated using the APBS plug-in⁴⁰ from PyMOL.



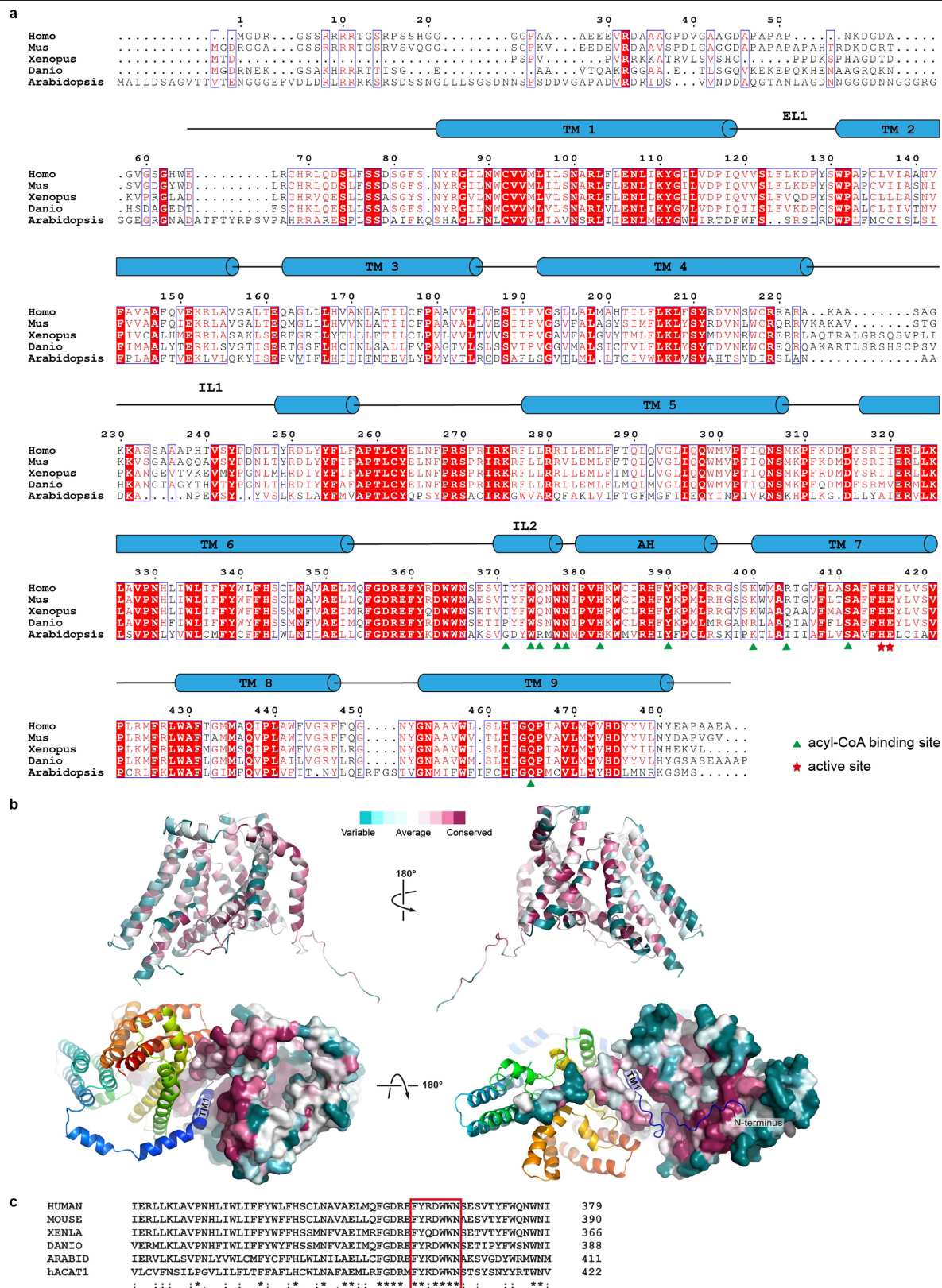
Extended Data Fig. 5 | Binding of the N-terminus to the neighbouring protomer and its functional consequences. a, b, The DGAT1 dimer (cartoon) viewed in two orientations. **c–e**, Detailed view of the boxed regions shown in **a, b**.

Residues involved in the interactions are shown as sticks. **f**, Size-exclusion profiles of N-terminal truncations. Experiments in **f** were repeated independently three times with similar results.



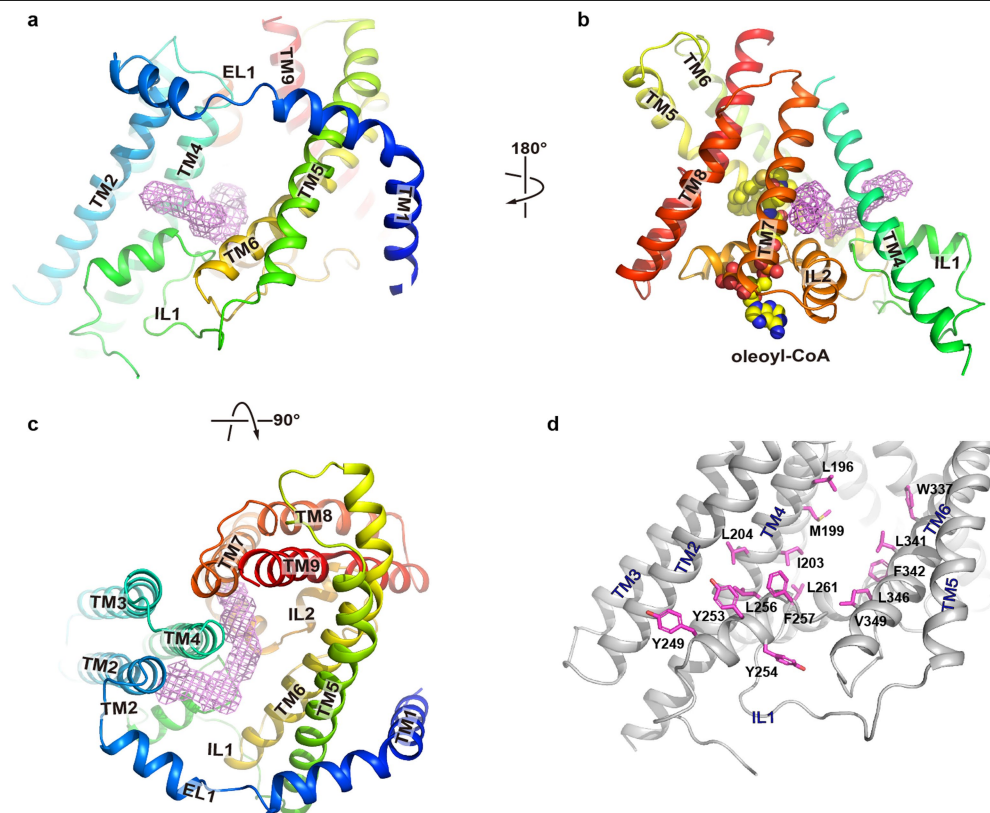
Extended Data Fig. 6 | Oleoyl-CoA-binding site. **a–c**, Oleoyl-CoA (spheres) bound to the DGAT1 protomer (cartoon) in three orientations. **d–g**, Detailed views of the boxed regions in **a–c**. Residues that coordinate oleoyl-CoA are shown as sticks with carbon atoms in magenta. **h**, LigPlus^{41,42} plot of the

oleoyl-CoA-binding site. **i**, Normalized enzymatic activity of wild-type DGAT1 and various mutants. Data are mean \pm s.e.m. derived from three independent repeats. **j**, Size-exclusion profiles of DGAT1 mutants. Experiments in **i** were repeated independently three times with similar results.



Extended Data Fig. 7 | DGAT1 sequence alignment. **a**, DGAT1 from human (*Homo*, Uniprot accession number O75907), mouse (*Mus*, Q9Z2A7), frog (*Xenopus*, A0A1L8G0L4), fish (*Danio*, Q6P3J0) and thale cress (*Arabidopsis*, Q9SLD2) are aligned using the Clustal Omega server⁴³. Secondary structural elements of human DGAT1 are marked above the alignment. Residues are coloured on the basis of their conservation using the ESPript server⁴⁴. Residues at the acyl-CoA-binding site are marked with green triangles and those at the

active site with red stars. **b**, ConSurf mapped onto the human DGAT1 structure. DGAT1 is shown in different orientations in cartoon (top) and surface (bottom) representations. One DGAT1 monomer is coloured on the basis of the conservation score of each residue, calculated by the ConSurf server⁴⁵. **c**, Sequence alignments of the FYXDWNN motif (red rectangle) from DGAT1 of different species and human ACAT1 (Uniprot P35610).



Extended Data Fig. 8 | Proposed gateway for DAG entry. **a–c**, A large opening to the core of the membrane is framed by TM4 and TM6. A tubular density is observed extending into the reaction chamber. **d**, Residues that line the opening are shown as magenta sticks.

Extended Data Table 1 | Summary of cryo-EM data collection, processing and structure refinement

	hDGAT1 (EMDB-21302) (PDB 6VP0)
Data collection and processing	
Magnification	105,000
Voltage (kV)	300
Electron exposure (e-/Å ²)	50
Defocus range (μm)	[-2.0, -1.2]
Pixel size (Å)	1.114
Symmetry imposed	C2
Initial particle images (no.)	2,749,110
Final particle images (no.)	275,945
Map resolution (Å)	3.1
FSC threshold	0.143
Map resolution range (Å)	2.7-4.3
Refinement	
Initial model used (PDB code)	6VP0
Model resolution (Å)	3.24
FSC threshold	0.5
Model resolution range (Å)	3.24-3.33
Map sharpening <i>B</i> factor (Å ²)	-100
Model composition	
Non-hydrogen atoms	7212
Protein residues	808
Ligands	12
<i>B</i> factors (Å ²)	
Protein	70.5
Ligand	78.1
R.m.s. deviations	
Bond lengths (Å)	0.005
Bond angles (°)	1.129
Validation	
MolProbity score	1.69
Clashscore	4.65
Poor rotamers (%)	1.96
Ramachandran plot	
Favored (%)	96.5
Allowed (%)	3.5
Disallowed (%)	0

Extended Data Table 2 | Functional parameters of wild-type human DGAT1 and various DGAT1 mutants

substrates/mutants/truncations	K _M (μM)	V _{max} (pmol/min/μg)
oleoyl-CoA	14.6 ± 1.3	956.6 ± 36.1
stearoyl-CoA	8.6 ± 1.3	839.4 ± 49.9
palmitoleoyl-CoA	6.2 ± 0.9	838.6 ± 31.6
palmitoyl-CoA	6.4 ± 1.1	767.8 ± 34.0
octodecyl-CoA	10.5 ± 1.4	642.9 ± 25.0
1,2-dioleoyl-sn-glycerol	597.1 ± 94.5	3310 ± 279.1 (h=1.50)
1,2-dioleoyl-sn-glycerol (tetramer)	497.5 ± 57.6	3628 ± 187.4 (h=1.05)
WT (dimer)	14.6 ± 1.3	956.6 ± 36.1
WT (tetramer)	16.6 ± 2.2	1080.8 ± 45.3
WT (membrane)	15.9 ± 1.3	1643.4 ± 36.4
WT (membrane)	15.9 ± 1.3	1643.4 ± 36.4
ΔN65	13.9 ± 2.6	563.9 ± 32.5
ΔN70	24.4 ± 1.5	540.7 ± 25.5
ΔN75	28.2 ± 3.1	497.8 ± 21.9
ΔN80	28.6 ± 4.2	276.1 ± 16.4
ΔN84	*	*

K_m and V_{max} values were obtained from fitting the initial rate versus concentration plots shown in Fig. 1, Extended Data Fig. 2 with the equations that are defined in Methods. The errors are 95% confidence intervals. For each concentration of substrate, the rate was measured three times independently.

Reporting Summary

Nature Research wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Research policies, see [Authors & Referees](#) and the [Editorial Policy Checklist](#).

Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

n/a Confirmed

- ☐ ☒ The exact sample size (n) for each experimental group/condition, given as a discrete number and unit of measurement
- ☐ ☒ A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
- ☐ ☒ The statistical test(s) used AND whether they are one- or two-sided
Only common tests should be described solely by name; describe more complex techniques in the Methods section.
- ☒ ☐ A description of all covariates tested
- ☒ ☐ A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
- ☐ ☒ A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)
- ☒ ☐ For null hypothesis testing, the test statistic (e.g. F , t , r) with confidence intervals, effect sizes, degrees of freedom and P value noted
Give P values as exact values whenever suitable.
- ☒ ☐ For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
- ☒ ☐ For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
- ☒ ☐ Estimates of effect sizes (e.g. Cohen's d , Pearson's r), indicating how they were calculated

Our web collection on [statistics for biologists](#) contains articles on many of the points above.

Software and code

Policy information about [availability of computer code](#)

Data collection

Serial EM 3.7 is used to collect cryo-EM data.

Data analysis

MotionCor2 1.1.0; GCTF 1.06; RELION 2.1; RELION-3.0-beta; Chimera 1.13; Coot 0.8.6.1; Phenix 1.13; Pymol 1.8.6.0; Prism 8.2.1; CryoSPARC 2.10.1.

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors/reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Research [guidelines for submitting code & software](#) for further information.

Data

Policy information about [availability of data](#)

All manuscripts must include a [data availability statement](#). This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A list of figures that have associated raw data
- A description of any restrictions on data availability

PDB coordinates has been deposited into the PDB database (6VP0). EM map has been deposited into the EMDB database (EMD-21302). All data will be available upon publication.

Field-specific reporting

Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

☒ Life sciences ☐ Behavioural & social sciences ☐ Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see [nature.com/documents/nr-reporting-summary-flat.pdf](https://www.nature.com/documents/nr-reporting-summary-flat.pdf)

Life sciences study design

All studies must disclose on these points even when the disclosure is negative.

Sample size	Sample sizes were chosen to give s.e.m. values of less than 10% of peak values based on prior experimental experience.
Data exclusions	No data were excluded from the analyses.
Replication	Each data point in enzymatic assay was measured in at least 3 independent experiments.
Randomization	No randomization is needed for enzymatic assay.
Blinding	All the mutants were tested and analyzed blindly to avoid bias.

Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

Materials & experimental systems

n/a	Involved in the study
<input checked="" type="checkbox"/>	<input type="checkbox"/> Antibodies
<input type="checkbox"/>	<input checked="" type="checkbox"/> Eukaryotic cell lines
<input checked="" type="checkbox"/>	<input type="checkbox"/> Palaeontology
<input checked="" type="checkbox"/>	<input type="checkbox"/> Animals and other organisms
<input checked="" type="checkbox"/>	<input type="checkbox"/> Human research participants
<input checked="" type="checkbox"/>	<input type="checkbox"/> Clinical data

Methods

n/a	Involved in the study
<input checked="" type="checkbox"/>	<input type="checkbox"/> ChIP-seq
<input checked="" type="checkbox"/>	<input type="checkbox"/> Flow cytometry
<input checked="" type="checkbox"/>	<input type="checkbox"/> MRI-based neuroimaging

Eukaryotic cell lines

Policy information about [cell lines](#)

Cell line source(s)	High Five Cells (invitrogen)
Authentication	Authentication No further authentication was performed for commercially available cell lines.
Mycoplasma contamination	The cell lines were not tested for mycoplasma contamination.
Commonly misidentified lines (See ICLAC register)	No commonly misidentified lines were used.

Structural basis for catalysis and substrate specificity of human ACAT1

<https://doi.org/10.1038/s41586-020-2290-0>

Received: 24 May 2019

Accepted: 17 March 2020

Published online: 13 May 2020

 Check for updates

Hongwu Qian^{1,7}✉, Xin Zhao^{2,7}, Renhong Yan^{3,4,7}, Xia Yao¹, Shuai Gao¹, Xue Sun⁵, Ximing Du⁶, Hongyuan Yang⁶, Catherine C. L. Wong⁵ & Nieng Yan¹✉

As members of the membrane-bound *O*-acyltransferase (MBOAT) enzyme family, acyl-coenzyme A:cholesterol acyltransferases (ACATs) catalyse the transfer of an acyl group from acyl-coenzyme A to cholesterol to generate cholesteryl ester, the primary form in which cholesterol is stored in cells and transported in plasma¹. ACATs have gained attention as potential drug targets for the treatment of diseases such as atherosclerosis, Alzheimer's disease and cancer^{2–7}. Here we present the cryo-electron microscopy structure of human ACAT1 as a dimer of dimers. Each protomer consists of nine transmembrane segments, which enclose a cytosolic tunnel and a transmembrane tunnel that converge at the predicted catalytic site. Evidence from structure-guided mutational analyses suggests that acyl-coenzyme A enters the active site through the cytosolic tunnel, whereas cholesterol may enter from the side through the transmembrane tunnel. This structural and biochemical characterization helps to rationalize the preference of ACAT1 for unsaturated acyl chains, and provides insight into the catalytic mechanism of enzymes within the MBOAT family⁸.

The synthesis of cholesteryl ester from acyl-coenzyme A (acyl-CoA) and cholesterol is critical for the formation of chylomicron in enterocytes, very low density lipoprotein in hepatocytes and lipid droplets in macrophages⁹. In addition to the ACAT enzymes that catalyse this transformation, other members of the MBOAT family include key enzymes in lipid metabolism—such as acyl-coenzyme A:diacylglycerol acyltransferase (DGAT) and lysophospholipid acyltransferases—and several protein acyltransferases that engage in signalling, such as ghrelin *O*-acyltransferase (GOAT), Wnt acyltransferase (Porcupine) and Hedgehog acyltransferase^{8,10–14}.

Two mammalian isoforms of ACAT have been identified, ACAT1 and ACAT2, and they share around 47% sequence identity^{15,16} (Extended Data Fig. 1). ACAT1 is mainly expressed in the liver, adrenal glands, macrophages and kidneys, whereas ACAT2 is specifically expressed in the intestines and the liver¹⁷. ACAT1 is an endoplasmic reticulum membrane protein that consists of nine predicted transmembrane segments (TMs)¹⁸ and an amino-terminal cytosolic domain (NTD) that is responsible for tetramerization¹⁹. The key active-site residue of ACAT1, His460, is conserved among members of the MBOAT family and is predicted to be on TM7^{8,18,20}. Some steroid molecules—such as sitosterol, cholestanol, allocholesterol and progesterone, all of which contain a 3β-OH group—are also substrates for ACAT1^{21,22}. An additional cholesterol-binding site is suggested to be involved in allosteric activation of the enzyme^{21–23}.

Accumulation of cholesteryl ester in lipid droplets is a main characteristic of macrophage foaming, which can lead to atherosclerotic diseases²⁴. Inhibition of ACATs may therefore represent a promising

strategy for the treatment of atherosclerosis, although results from clinical trials have so far been inconclusive^{2,25–28}. In addition, using mouse models, inhibitors of ACAT1 have previously been shown to alleviate amyloid pathology²⁹, reduce the size of hepatocellular carcinoma tumours⁴, suppress the growth and metastasis of pancreatic cancer tumours⁵, prevent prostate cancer⁶ and enhance the antitumour response of CD8⁺ T cells and immunotherapy⁷.

A lack of structural information regarding ACATs has impeded understanding of their mechanism of action and research into drugs that target them. The only available structure of an enzyme in the MBOAT family is that of a bacterial homologue, DltB, which shares low sequence similarity with eukaryotic ACATs³⁰. Here we report the structure of full-length human ACAT1 at an overall resolution of 3.1 Å for the dimer of dimers and 3.0 Å for the dimer only. We have also established an *in vitro* assay to measure the activity of ACAT variants. Structure-guided biochemical and mass spectrometric characterization enable us to suggest the substrate entry and product exit pathways, and elucidate the molecular basis for the substrate preference of ACAT1.

Structural determination of ACAT1

The recombinant expression and purification of ACAT1 is described in detail in the Methods (Fig. 1a). To examine the enzymatic activity of the purified protein, we developed a fluorescence-based assay (Extended Data Fig. 2a). The purified protein was mixed with cholesterol-containing 1-palmitoyl-2-oleoyl phosphatidylcholine (POPC) micelles according to a reported protocol^{31,32}, and the acyl

¹Department of Molecular Biology, Princeton University, Princeton, NJ, USA. ²State Key Laboratory of Membrane Biology, Beijing Advanced Innovation Center for Structural Biology, Tsinghua-Peking Joint Center for Life Sciences, School of Life Sciences, Tsinghua University, Beijing, China. ³Key Laboratory of Structural Biology of Zhejiang Province, School of Life Sciences, Westlake University, Hangzhou, China. ⁴Institute of Biology, Westlake Institute for Advanced Study, Hangzhou, China. ⁵Center for Precision Medicine Multi-omics Research, State Key Laboratory of Natural and Biomimetic Drugs, School of Pharmaceutical Sciences, Peking-Tsinghua Center for Life Sciences, Peking University, Beijing, China. ⁶School of Biotechnology and Biomolecular Science, The University of New South Wales, Sydney, New South Wales, Australia. ⁷These authors contributed equally: Hongwu Qian, Xin Zhao, Renhong Yan. ✉e-mail: hongwuq@princeton.edu; nyan@princeton.edu

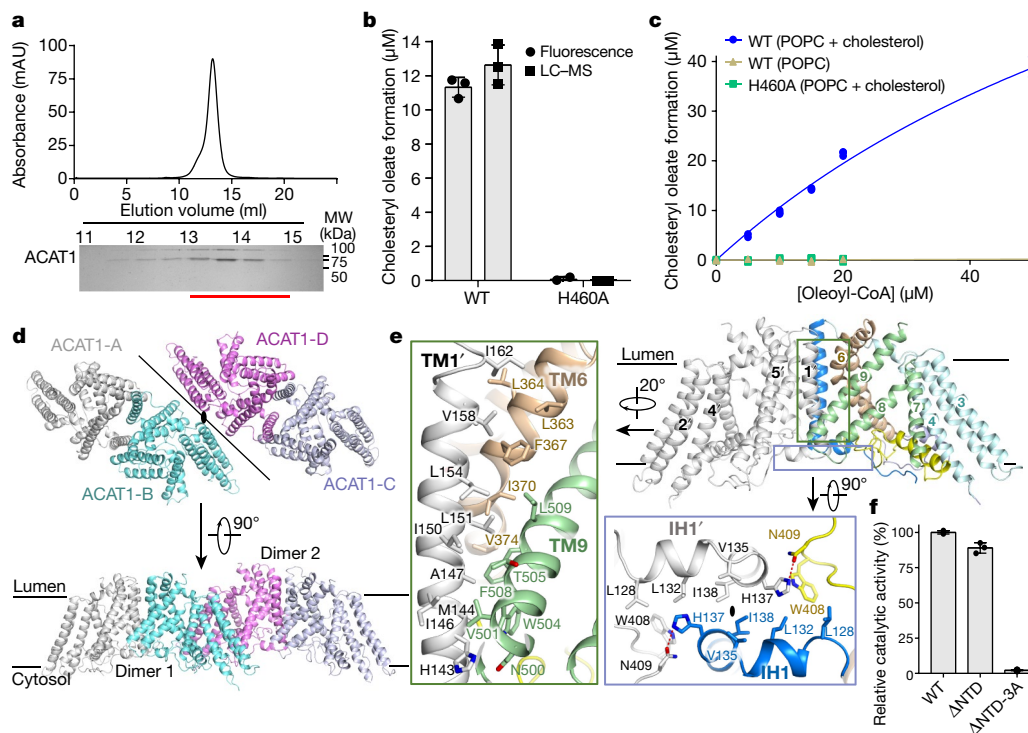


Fig. 1 | Cryo-EM structure of human ACAT1 as a dimer of dimers. **a**, Trace from the SEC purification of ACAT1 using Superose 6. The highlighted fractions (13–15 ml) on the SDS–PAGE gel were concentrated for cryo-sample preparation. The experiments were independently repeated three times with similar results. **b**, Activity of wild-type (WT) ACAT1 and the ACAT1(H460A) variant, as measured using the fluorescence-based assay and LC–MS. **c**, Formation of cholesteryl oleate in 10 minutes catalysed by wild-type ACAT1 in the presence and absence of cholesterol, and by ACAT1(H460A), measured

using the fluorescence-based assay. **d**, ACAT1 exists as dimer of dimers. The two protomers that mediate the interface of dimers, indicated by the black line, are coloured cyan and magenta. The black oval indicates the C2 symmetric axis. **e**, The dimeric interface of ACAT1. Protomer A is coloured grey and protomer B is domain-coloured. The potential hydrogen bonds are represented by dashed red lines. **f**, ACAT1 is functional as a dimer (ACAT1(ΔNTD)), but is almost inactive as a monomer (ACAT1(ΔNTD-3A)). Data in **b**, **c**, **f** are mean \pm s.d. of three independent experiments.

transfer reaction was initiated by the addition of oleoyl-CoA. The release of free CoA (CoASH) was detected using the fluorescent probe 7-diethylamino-3-(4-maleimidophenyl)-4-methylcoumarin³³. The catalytic activity measured in the fluorescence-based assay was validated using liquid chromatography coupled to mass spectrometry (LC–MS) analysis, which enabled detection of the second product, cholesteryl oleate (Fig. 1b).

ACAT1 purified in 1% (w/v) CHAPS gave a better signal-to-noise ratio in the fluorescence assay than that purified in 0.02% (w/v) glyco-diosgenin (GDN) (Extended Data Fig. 2b). Therefore, wild-type ACAT1 and ACAT1 mutants were examined in the presence of CHAPS in all assays described hereafter. No fluorescence signal was detected in the absence of cholesterol, confirming that acyl-CoA was not hydrolysed under these conditions. With increasing concentrations of cholesterol the assay data were fit to a sigmoidal curve, consistent with previous reports that ACAT1 is allosterically activated by cholesterol³². A single point mutation of the histidine at position 460 to alanine resulted in the complete abolition of enzymatic activity^{18,20} (Fig. 1b, c, Extended Data Fig. 2c).

ACAT1 has previously been found to exhibit a preference for oleoyl-CoA over stearoyl-CoA as the substrate³². To elucidate the molecular basis for catalysis and for this substrate specificity, we set out to determine the structure of ACAT1. Protein purified in 0.02% GDN yielded homogenous particles that were suitable for analysis by cryo-electron microscopy (cryo-EM) (Extended Data Figs. 3, 4, Extended Data Table 1). Representative 2D averages indicated an assembly of dimer of dimers (Extended Data Fig. 3a), consistent with the previously reported tetrameric organization of ACAT1³⁴. Eventually, 183,794 selected particles yielded a 3D reconstruction with an overall resolution of 3.1 Å for the tetramer (Extended Data Fig. 3b, c). To further

improve the resolution, we expanded the particles with C2 symmetry and focused on one dimer with an adapted mask for refinement. The resolution was improved to 3.0 Å after further 3D classification (Extended Data Figs. 3b–d, 4a, Extended Data Table 1), and an atomic model was built on the basis of the dimeric reconstruction (Fig. 1d, Extended Data Figs. 3e, 4b–d).

In each protomer, 408 residues (118–281 and 286–529) could be resolved, of which 405 side chains were assigned (Extended Data Table 1). In addition to the predicted nine TMs¹⁸, three helices were resolved on the intracellular side—designated IH1, IH2 and IH3—and one on the extracellular side that was termed EH1. Whereas TM1 in each protomer extends out to enable dimerization, TM2–TM9 fold into three sub-domains: TM2–TM4, TM5–TM6 and TM7–TM9. These subdomains are interspersed by two elongated intracellular loops, Loop1 and Loop2, each of which carries an intracellular helix (IH2 and IH3, respectively) (Extended Data Fig. 4d). The overall transmembrane structure of each ACAT1 protomer contains several tunnels and cavities, which we describe in detail.

Dimer of dimers

In the tetrameric ACAT1, the two dimers make only limited contact with each other within the membrane itself, through an interface that involves TM2, TM5, TM6 and IH2 from the two protomers in the centre (Fig. 1d, Extended Data Fig. 5a, b). This limited transmembrane interface supports previous findings that two α -helices in the NTD are required for tetramerization¹⁹.

Corroborating this notion, densities that probably belong to the NTD are observed in a cluster on the cytosolic side when the map is

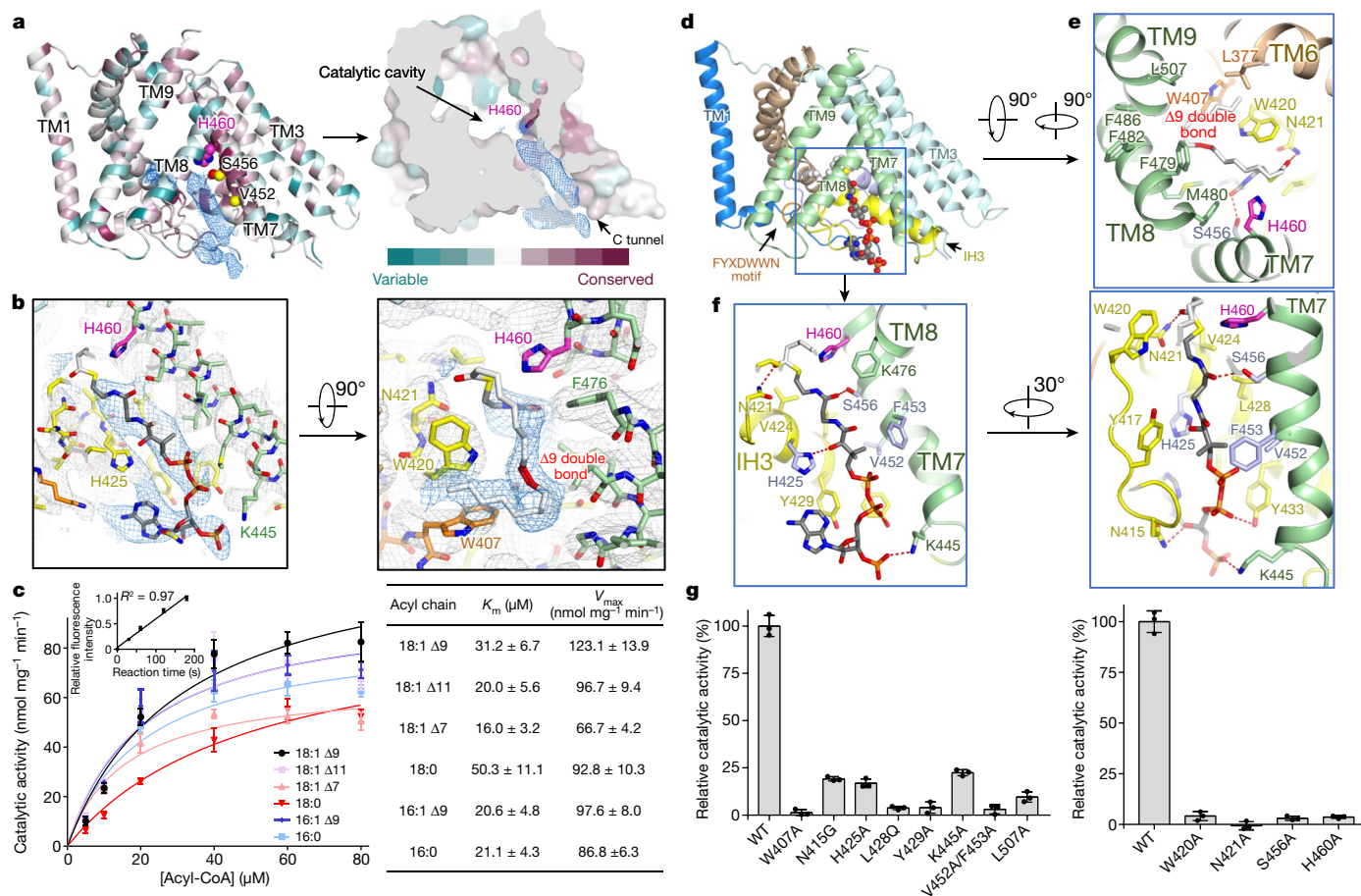


Fig. 2 | The C tunnel in the ACAT1 protomer may accommodate an oleoyl-CoA molecule. **a**, An elongated density that is reminiscent of oleoyl-CoA is observed in the highly conserved C tunnel in each protomer. The density, shown as blue mesh, is contoured at 6σ . See Extended Data Fig. 7 for mass spectrometric analysis. **b**, Densities of the head group (left) and aliphatic tail (right) of oleoyl-CoA. Densities, contoured at 6σ , are from ACAT1-B. **c**, Substrate preference of ACAT1. Left, enzymatic activity in the presence of substrates with different acyl chains. The inset shows that the fluorescence signal is almost linear during the first 200 s of the reaction. Right, summary of the Michaelis constant (K_m) and maximum rate of reaction (V_{max}) of the enzyme with the tested substrates. **d**, The oleoyl-CoA binding tunnel is mainly

surrounded by TM7, TM8 and IH3. The FYXDWWN motif is coloured orange. **e**, Potential coordination of the oleoyl tail in the catalytic cavity. The contour of the cavity supports the finding that unsaturated acyl chains are preferred as substrates over the corresponding analogues with saturated chains. The double bond at $\Delta 9$ of the oleoyl tail is coloured red. **f**, Potential coordination of the head group of oleoyl-CoA by ACAT1. The residues for which biochemical characterization has previously been reported^{20,40} are coloured light purple. **g**, Functional verification of the importance of the residues that engage in oleoyl-CoA coordination. The activities of the mutants were normalized relative to that of the wild type. Data in **c**, **g** are mean \pm s.d. of three independent experiments.

low-pass filtered (Extended Data Fig. 5a). We generated a mutant in which the N-terminal 109 residues were truncated (ACAT1(Δ NTD)). During purification by size-exclusion chromatography (SEC) this variant eluted around 1.0 ml later than wild-type ACAT1, indicating disruption of the tetramer (Extended Data Fig. 5c). To further investigate the oligomeric state we acquired cryo-EM images of ACAT1(Δ NTD), and 2D class averages showed reduced particle sizes consistent with a dimer (Extended Data Fig. 5d). These results support a role for the NTD in the dimerization of dimers.

The two protomers in each dimer, which exhibit an approximate C2 symmetry around an axis perpendicular to the membrane plane, can be superimposed with a root mean square deviation (r.m.s.d.) of approximately 0.96 Å over 407 C α atoms (Extended Data Fig. 5e, f). The dimerization of ACAT1 is mainly mediated by extensive van der Waals interactions between TM1 in one protomer and the luminal segment of TM6 and the cytosolic segment of TM9 in the other; in general, each TM1 segment packs predominantly with the remaining TM domains of the opposite protomer (Fig. 1e). TM1, TM5, TM6 and TM9 from the two protomers enclose a deep hydrophobic pocket that is open to the luminal side (Extended Data Fig. 5g). Numerous hydrophobic residues

on TM6 and TM9 from one protomer contact those on TM1 from the other protomer (labelled as TM1'). On the intracellular side, hydrophobic residues on IH1 of each protomer interact with each other to stabilize the dimer. In addition, His137 seems to be hydrogen-bonded to Asn409' (Fig. 1e).

To confirm the interactions involved in the dimeric interface, three aromatic residues of ACAT1(Δ NTD)—Phe367, Trp504 and Phe508—were substituted for alanine to form a mutant that we denote ACAT(Δ NTD-3A). This mutant eluted approximately 1 ml after ACAT1(Δ NTD) during SEC, indicating disruption of the dimer (Extended Data Fig. 5c). The catalytic activities of tetrameric, dimeric and monomeric ACAT1 were measured using the fluorescence assay. The activities of the tetramer and the dimer were comparable, whereas that of the monomer was undetectable (Fig. 1f), suggesting that the dimer is the catalytic unit.

Entrance for acyl-CoA

In each ACAT protomer, the conserved catalytic residue His460 is located in the middle of the structure, which indicates that catalysis

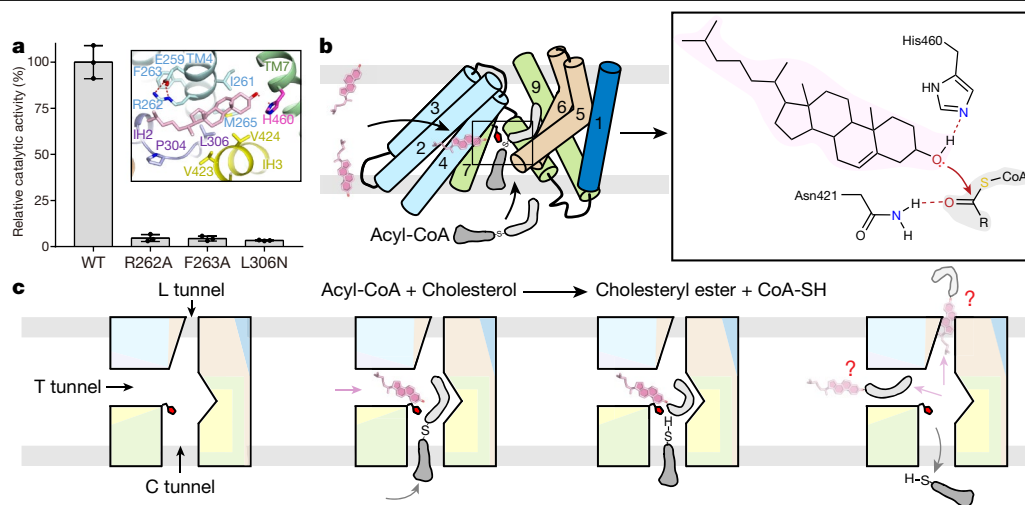


Fig. 3 | Structure-based working model for ACAT1. a, The residues that constitute the T tunnel are critical to the activity of ACAT1. Data are mean \pm s.d. of three independent experiments. Inset, a cholesterol molecule docked into the T tunnel using the software Schrödinger. The T tunnel may serve as the entrance for cholesterol. **b**, Putative mechanism for ACAT1-catalysed acyl transfer. **c**, Working model of ACAT1. The T tunnel and the C tunnel may provide

an entrance for cholesterol and acyl-CoA, respectively. The reaction is catalysed at the intersection of the two tunnels, where the active residue His460 is located. Regarding exit of the products, CoASH can be released to the cytosol via the C tunnel, whereas cholesterol ester may exit either through the T tunnel to the membrane or through the L tunnel to the lumen, a question that remains to be investigated.

potentially occurs within the membrane. TM2–TM9 enclose a hydrophilic environment for the acyl-transfer reaction. TM4 along with TM6–TM9 constitute a catalytic cavity that opens to the cytosolic side through an elongated tunnel, which is designated the cytosolic (C) tunnel. The residues that line the C tunnel are conserved among members of the MBOAT family (Fig. 2a), and TM2–TM9 of ACAT1 and TM3–TM10 of DltB can be superimposed with an r.m.s.d. of 5.0 Å over 272 C α atoms, indicating that MBOAT enzymes feature a conserved catalytic domain³⁰ (Extended Data Fig. 6a, b). However, the C tunnel is missing in DltB because of the conformational shifts of TM9, which corresponds to TM8 in ACAT1 (Extended Data Fig. 6a, c).

The bottom of the catalytic cavity in ACAT1 is formed mainly by the conserved Loop1 and Loop2, and the top opens to the luminal side (Extended Data Figs. 1, 6b, c). An elongated density was observed in the C tunnel of each protomer (Fig. 2a). Because no substrates were added during protein purification or cryo-sample preparation, this density may belong to an endogenous ligand.

Oleoyl-CoA can be well fitted into the active-site density (Fig. 2b, Extended Data Fig. 7a, b). To confirm this, lipids in purified ACAT1 were extracted and subject to LC–MS analysis. A peak corresponding to the positive control—a commercial sample of oleoyl-CoA that represents the preferred form of acyl-CoA for ACAT1³²—appeared in the liquid chromatography profile (Extended Data Fig. 7c). The tandem mass spectrometry (MS/MS) profile of this peak was consistent with the MS/MS fragmentation pattern of oleoyl-CoA, confirming the presence of oleoyl-CoA in the purified enzymes (Extended Data Fig. 7d, e).

On the basis of structural analysis, we generated the double mutant ACAT1(V452Q/S456Q)—designated the QQ mutant—in which the C tunnel was expected to be obstructed (Fig. 2a). This mutant showed very little enzymatic activity (Extended Data Figs. 7f, 8a). LC–MS analysis of extracts from experiments with the QQ mutant showed a marked decrease of the oleoyl-CoA peak (Extended Data Fig. 7c, d), further supporting the presence of oleoyl-CoA in the purified ACAT1.

The CoA head group was found to fit reasonably into the density (Fig. 2b). Initially we did not observe any corresponding density for the oleoyl tail contiguous with the head group in the catalytic cavity; however, we subsequently identified a stretch of density close to the CoA that could belong to the oleoyl tail, despite its disconnection from the head density. The oleoyl tail, bent at the position of the $\Delta 9$ double

bond, could be perfectly docked into this density (Fig. 2b). Supporting this observation, ACAT1 showed a higher activity towards an acyl-CoA substrate with a $\Delta 9$ double bond than towards saturated acyl-CoA or an unsaturated acyl-CoA with a $\Delta 7$ or $\Delta 11$ double bond (Fig. 2c).

The oleoyl tail is surrounded by several hydrophobic residues on TM6, TM8, TM9 and Loop2 (Fig. 2d, e). Substitution of the spatially adjacent Trp407 or Trp420 on Loop2, or Leu507 on TM9, with alanine markedly reduced the catalytic activity (Fig. 2g, Extended Data Fig. 8a). Trp407 is part of the conserved FYXDWWN motif, which has been suggested to participate in acyl-CoA coordination³⁵ (Fig. 2d, e, Extended Data Fig. 1a).

The CoA group is enclosed by TM7, TM8, IH3 and the loop between TM6 and IH3 (Fig. 2d, f). The adenine moiety of oleoyl-CoA may be stabilized through π – π stacking with Tyr429, the substitution of which by alanine led to an approximately 90% reduction in activity (Fig. 2f, g, Extended Data Fig. 8a). The phosphate seems to be anchored through a salt bridge with Lys445, in addition to hydrogen bonds with Tyr433 and Asn415. Supporting this observation, the mutation of Lys445 or of Asn415 led to a decrease in enzymatic activity (Fig. 2g, Extended Data Fig. 8a). The elongated pantetheine arm, which may be stabilized through hydrogen bonds with His425 and Ser456, penetrates the tunnel to place the thioester bond adjacent to His460 (Fig. 2f). Consistent with this, the activity of the ACAT1(H425A) mutant was reduced by up to 80% compared with the wild type, and the ACAT1(S456A) mutant showed only residual activity, as has previously been reported²⁰ (Fig. 2g, Extended Data Fig. 8a).

In addition to the polar interactions, several non-polar residues—including Val424, Leu428, Val452 and Phe453—stabilize the pantetheine group through van der Waals contacts. Mutational analysis revealed that the activities of the ACAT1(V452A/F453A) and ACAT1(L428Q) mutants were only around 5% that of the wild type (Fig. 2f, g, Extended Data Fig. 8a).

A transmembrane tunnel to the active site

The structure of ACAT1 features a tunnel—enclosed by EH1, TM2, TM4–6 and Loop 1—that traverses the middle of the structure and opens to the centre of the membrane. We designate this as the transverse or transmembrane (T) tunnel (Extended Data Fig. 9a). Notably, the T tunnel

and the aforementioned C tunnel converge at the catalytic residue His460. An elongated density, different in shape from cholesterol, is observed in the T tunnel (Extended Data Fig. 9b). Single point mutation of the tunnel residues to generate ACAT1(R262A), ACAT1(F263A) and ACAT1(L306N) resulted in a greater than 90% reduction in activity compared with the wild type (Fig. 3a, Extended Data Fig. 8b), suggesting that the T tunnel has a critical role in catalysis.

Discussion

Enzymes of the MBOAT family, which catalyse the transfer of an acyl group to receptors including lipids and proteins, have critical roles in lipid metabolism and signal transduction. On the basis of our structural observations and biochemical analysis, we propose a catalytic model for ACAT1 that will also shed light on the mechanistic understanding of other MBOAT enzymes (Fig. 3b, c).

We suggest that the C tunnel and the T tunnel, which converge at the catalytic residue His460, represent the entrances for the two substrates—the C tunnel seems to enable the access of acyl-CoA, whereas T tunnel is likely to be responsible for cholesterol entry. Although cholesterol cannot be fit into the density in the T tunnel, simple docking suggests that the tunnel can accommodate a cholesterol molecule with the 3 β -OH pointing towards His460 (Fig. 3a). The histidine residue that is conserved at the active site of several transferases—such as carnitine acetyltransferase, cholesterol sulfotransferase and UDP-N-acetylglucosamine acyltransferase^{36–38}—can function as a general base to deprotonate and activate the nucleophilic substrate (Extended Data Fig. 9c). For ACAT1-catalysed acyl transfer, the 3 β -OH of cholesterol may act as the nucleophilic substrate that is activated through deprotonation, most likely by His460.

In our structure, the thioester bond of oleoyl-CoA is surrounded by two conserved residues—Asn421 and His460—which are on opposite sides of the thioester of oleoyl-CoA (Fig. 2e, f, Extended Data Fig. 1a). His460, which is highly conserved among enzymes of the MBOAT family, was previously shown to be essential to the catalytic activity¹⁸. The highly conserved residue Asn421 has been suggested to act as a second active site¹⁴. Supporting the importance of Asn421, no activity was detected for the ACAT1(N421A) mutant (Fig. 2g, Extended Data Fig. 8a).

Asn421 may form a hydrogen bond with the carbonyl oxygen of oleoyl-CoA to facilitate the nucleophilic attack (Fig. 3b). After the formation and exit of cholesteryl ester, His460 is deprotonated and the proton is transferred to the sulfur atom to form the other product, CoASH, which can be conveniently released into the cytosol through the C tunnel (Fig. 3c). It remains to be determined whether cholesteryl ester diffuses to the membrane via the T tunnel, or whether it is taken—by as-yet-uncharacterized receptors—to the lumen via the luminal (L) tunnel (Fig. 3c).

A conserved Ser–His–Asp catalytic triad—involving Ser456, His460 and Asp400—has been suggested to participate in catalysis²⁰. However, the structure shows that Asp400 is far away from the active site, and so is unlikely to be a component of the catalytic triad. Ser456, which is one helical turn below His460, may help to stabilize acyl-CoA but is not involved in catalysis (Fig. 2f).

Several different acyl-CoA molecules can act as acyl donors for ACAT1, among which unsaturated acyl-CoAs are preferred over their corresponding saturated analogues. The contour of the substrate-accommodating tunnel and the results of enzymatic assays suggest that, among the unsaturated acyl-CoAs, a double bond in the Δ 9 position may be the most favourable (Fig. 2b, c). Supporting this analysis, mice with a mutation in stearoyl-CoA desaturase 1 (also known as Δ 9 desaturase)—which introduces a *cis*-double bond at the Δ 9 position of predominantly stearoyl- and palmitoyl-CoA—were deficient in cholesteryl ester biosynthesis in hepatocytes³⁹.

In terms of future research, one major question awaits further investigation: the location and nature of the cholesterol-binding site

for allosteric activation^{21,22}. However, although our knowledge of this enzyme is not yet complete, the cryo-EM structure of human ACAT1 and the fluorescence-based assay that we describe here should serve as a foundation for future studies and for drug discovery.

Online content

Any methods, additional references, Nature Research reporting summaries, source data, extended data, supplementary information, acknowledgements, peer review information; details of author contributions and competing interests; and statements of data and code availability are available at <https://doi.org/10.1038/s41586-020-2290-0>.

- Chang, T. Y., Chang, C. C. & Cheng, D. Acyl-coenzyme A:cholesterol acyltransferase. *Annu. Rev. Biochem.* **66**, 613–638 (1997).
- Ohshiro, T. et al. Pyripyropene A, an acyl-coenzyme A:cholesterol acyltransferase 2-selective inhibitor, attenuates hypercholesterolemia and atherosclerosis in murine models of hyperlipidemia. *Arterioscler. Thromb. Vasc. Biol.* **31**, 1108–1115 (2011).
- Hartmann, T., Kuchenbecker, J. & Grimm, M. O. Alzheimer's disease: the lipid connection. *J. Neurochem.* **103** (Suppl 1), 159–170 (2007).
- Jiang, Y. et al. Proteomics identifies new therapeutic targets of early-stage hepatocellular carcinoma. *Nature* **567**, 257–261 (2019).
- Li, J. et al. Abrogating cholesterol esterification suppresses growth and metastasis of pancreatic cancer. *Oncogene* **35**, 6378–6388 (2016).
- Yue, S. et al. Cholesteryl ester accumulation induced by PTEN loss and PI3K/AKT activation underlies human prostate cancer aggressiveness. *Cell Metab.* **19**, 393–406 (2014).
- Yang, W. et al. Potentiating the antitumour response of CD8⁺ T cells by modulating cholesterol metabolism. *Nature* **531**, 651–655 (2016).
- Hofmann, K. A superfamily of membrane-bound O-acyltransferases with implications for wnt signaling. *Trends Biochem. Sci.* **25**, 111–112 (2000).
- Rudel, L. L. & Shelleness, G. S. Cholesterol esters and atherosclerosis—a game of ACAT and mouse. *Nat. Med.* **6**, 1313–1314 (2000).
- Bhatt-Wessel, B., Jordan, T. W., Miller, J. H. & Peng, L. Role of DGAT enzymes in triacylglycerol metabolism. *Arch. Biochem. Biophys.* **655**, 1–11 (2018).
- Zhao, Y. et al. Identification and characterization of a major liver lysophosphatidylcholine acyltransferase. *J. Biol. Chem.* **283**, 8258–8265 (2008).
- Kojima, M., Hamamoto, A. & Sato, T. Ghrelin O-acyltransferase (GOAT), a specific enzyme that modifies ghrelin with a medium-chain fatty acid. *J. Biochem.* **160**, 189–194 (2016).
- Buglino, J. A. & Resh, M. D. Palmitoylation of Hedgehog proteins. *Vitam. Horm.* **88**, 229–252 (2012).
- Chang, C. C. Y., Sun, J. & Chang, T.-Y. Membrane-bound O-acyltransferases (MBOATs). *Front. Biol.* **6**, 177 (2011).
- Chang, C. C., Huh, H. Y., Cadigan, K. M. & Chang, T. Y. Molecular cloning and functional expression of human acyl-coenzyme A:cholesterol acyltransferase cDNA in mutant Chinese hamster ovary cells. *J. Biol. Chem.* **268**, 20747–20755 (1993).
- Cases, S. et al. ACAT-2, a second mammalian acyl-CoA:cholesterol acyltransferase. Its cloning, expression, and characterization. *J. Biol. Chem.* **273**, 26755–26764 (1998).
- Chang, C. C. et al. Immunological quantitation and localization of ACAT-1 and ACAT-2 in human liver and small intestine. *J. Biol. Chem.* **275**, 28083–28092 (2000).
- Guo, Z. Y., Lin, S., Heinen, J. A., Chang, C. C. & Chang, T. Y. The active site His-460 of human acyl-coenzyme A:cholesterol acyltransferase 1 resides in a hitherto undisclosed transmembrane domain. *J. Biol. Chem.* **280**, 37814–37826 (2005).
- Yu, C. et al. Role of the N-terminal hydrophilic domain of acyl-coenzyme A:cholesterol acyltransferase 1 on the enzyme's quaternary structure and catalytic efficiency. *Biochemistry* **41**, 3762–3769 (2002).
- Das, A., Davis, M. A. & Rudel, L. L. Identification of putative active site residues of ACAT enzymes. *J. Lipid Res.* **49**, 1770–1781 (2008).
- Liu, J., Chang, C. C., Westover, E. J., Covey, D. F. & Chang, T. Y. Investigating the allosterism of acyl-CoA:cholesterol acyltransferase (ACAT) by using various sterols: in vitro and intact cell studies. *Biochem. J.* **391**, 389–397 (2005).
- Rogers, M. A. et al. Cellular pregnenolone esterification by acyl-CoA:cholesterol acyltransferase. *J. Biol. Chem.* **287**, 17483–17492 (2012).
- Zhang, Y. et al. Cholesterol is superior to 7-ketocholesterol or 7 α -hydroxycholesterol as an allosteric activator for acyl-coenzyme A:cholesterol acyltransferase 1. *J. Biol. Chem.* **278**, 11642–11647 (2003).
- Ghosh, S., Zhao, B., Bie, J. & Song, J. Macrophage cholesteryl ester mobilization and atherosclerosis. *Vascul. Pharmacol.* **52**, 1–10 (2010).
- Tardif, J. C. et al. Effects of the acyl coenzyme A:cholesterol acyltransferase inhibitor avasimibe on human atherosclerotic lesions. *Circulation* **110**, 3372–3377 (2004).
- Chang, C. et al. Human acyl-CoA:cholesterol acyltransferase (ACAT) and its potential as a target for pharmaceutical intervention against atherosclerosis. *Acta Biochim. Biophys. Sin. (Shanghai)* **38**, 151–156 (2006).
- Nissen, S. E. et al. Effect of ACAT inhibition on the progression of coronary atherosclerosis. *N. Engl. J. Med.* **354**, 1253–1263 (2006).
- Meuwese, M. C. et al. ACAT inhibition and progression of carotid atherosclerosis in patients with familial hypercholesterolemia: the CAPTIVATE randomized trial. *J. Am. Med. Assoc.* **301**, 1131–1139 (2009).
- Shibuya, Y., Chang, C. C. & Chang, T. Y. ACAT1/SOAT1 as a therapeutic target for Alzheimer's disease. *Future Med. Chem.* **7**, 2451–2467 (2015).
- Ma, D. et al. Crystal structure of a membrane-bound O-acyltransferase. *Nature* **562**, 286–290 (2018).

31. Chang, C. C. et al. Recombinant acyl-CoA:cholesterol acyltransferase-1 (ACAT-1) purified to essential homogeneity utilizes cholesterol in mixed micelles or in vesicles in a highly cooperative manner. *J. Biol. Chem.* **273**, 35132–35141 (1998).
32. Chang, C. C. et al. Purification of recombinant acyl-coenzyme A:cholesterol acyltransferase 1 (ACAT1) from H293 cells and binding studies between the enzyme and substrates using difference intrinsic fluorescence spectroscopy. *Biochemistry* **49**, 9957–9963 (2010).
33. Cao, J. et al. Targeting acyl-CoA:diacylglycerol acyltransferase 1 (DGAT1) with small molecule inhibitors for the treatment of metabolic diseases. *J. Biol. Chem.* **286**, 41838–41851 (2011).
34. Yu, C. et al. Human acyl-CoA:cholesterol acyltransferase-1 is a homotetrameric enzyme in intact cells and in vitro. *J. Biol. Chem.* **274**, 36139–36145 (1999).
35. Guo, Z., Cromley, D., Billheimer, J. T. & Sturley, S. L. Identification of potential substrate-binding sites in yeast and human acyl-CoA sterol acyltransferases by mutagenesis of conserved sequences. *J. Lipid Res.* **42**, 1282–1291 (2001).
36. Hsiao, Y. S., Jogl, G. & Tong, L. Crystal structures of murine carnitine acetyltransferase in ternary complexes with its substrates. *J. Biol. Chem.* **281**, 28480–28487 (2006).
37. Lee, K. A. et al. Crystal structure of human cholesterol sulfotransferase (SULT2B1b) in the presence of pregnenolone and 3'-phosphoadenosine 5'-phosphate. Rationale for specificity differences between prototypical SULT2A1 and the SULT2BG1 isoforms. *J. Biol. Chem.* **278**, 44593–44599 (2003).
38. Ulaganathan, V., Buetow, L. & Hunter, W. N. Nucleotide substrate recognition by UDP-N-acetylglucosamine acyltransferase (LpxA) in the first step of lipid A biosynthesis. *J. Mol. Biol.* **369**, 305–312 (2007).
39. Miyazaki, M., Kim, Y. C., Gray-Keller, M. P., Attie, A. D. & Ntambi, J. M. The biosynthesis of hepatic cholesterol esters and triglycerides is impaired in mice with a disruption of the gene for stearoyl-CoA desaturase 1. *J. Biol. Chem.* **275**, 30132–30138 (2000).
40. Guo, Z. Y., Chang, C. C. & Chang, T. Y. Functionality of the seventh and eighth transmembrane domains of acyl-coenzyme A:cholesterol acyltransferase 1. *Biochemistry* **46**, 10063–10071 (2007).

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

© The Author(s), under exclusive licence to Springer Nature Limited 2020

Methods

Protein expression and purification

The cDNA of human ACAT1 (NCBI reference sequence NM_003101.6) was cloned into the pCAG vector with an amino-terminal Flag tag and a carboxy-terminal 10× His tag. All mutants were generated with a standard two-step PCR-based strategy. HEK293F suspension cells (Thermo Fisher Scientific, R79007) were cultured in Freestyle 293 medium (Thermo Fisher Scientific) at 37 °C, supplied with 5% CO₂ and 80% humidity. When cell density reached 2.0×10^6 cells per ml, the cells were transiently transfected with the expression plasmids and polyethylenimines (Polysciences). Approximately 1 mg of expression plasmids were pre-mixed with 3 mg polyethylenimines in 50 ml fresh medium and incubated for 15–30 min before transfection. The 50-ml mixture was then added to 1 l of cell culture and incubated for 15–30 min. Transfected cells were cultured for 48 h before collection.

For purification of ACAT1, the HEK293F cells were collected and resuspended in buffer containing 25 mM Tris pH 8.0, 150 mM NaCl and protease inhibitor cocktails (Amresco). After sonication on ice, the membrane fraction was solubilized at 4 °C for 2 h with 1% (w/v) GDN (Anatrace). After centrifugation at 20,000g for 1 h, the supernatant was collected and applied to anti-Flag M2 affinity resin (Sigma). The resin was rinsed with wash buffer (W1 buffer) containing 25 mM Tris pH 8.0, 150 mM NaCl and 0.02% GDN. The protein was eluted with W1 buffer plus 0.2 mg ml⁻¹ Flag peptide. The eluent was then concentrated and further purified by SEC (Superose 610/300 GL, GE Healthcare) in buffer containing 25 mM Tris pH 8.0, 150 mM NaCl and 0.02% GDN. The peak fractions were collected for cryo-sample preparation. The ACAT1 proteins for the fluorescence-based assay were purified similarly, except that the buffer for SEC was replaced by 25 mM HEPES pH 7.4, 150 mM NaCl and 1% CHAPS. After purification, a small amount of proteins was immediately subject to SEC to check their behaviour.

Fluorescence-based catalytic assay

To monitor the enzymatic activity of purified ACAT1 and its variants, 3 µl 0.8 mg ml⁻¹ protein was diluted into 37 µl solution containing micelles mixed with taurocholate, cholesterol and POPC, as described previously³¹, with the final concentration of taurocholate at 4 mM, POPC at 10 mM and cholesterol at 2 mM in HEPES buffer (25 mM HEPES 7.4, 150 mM NaCl). To measure the catalytic activity under different concentrations of cholesterol, the total amount of POPC and cholesterol was fixed at 12 mM. Another 3 µl of 1 M HEPES buffer (pH 7.4) was added to maintain a neutral environment. Acyl-CoA was added to initiate the reaction to the indicated concentration. The mutational analyses were performed in the presence of 40 µM oleoyl-CoA.

Unless specifically mentioned, all reactions were allowed to proceed for 3 min at room temperature. The reactions were stopped by the addition of SDS to a final concentration of 1% (w/v), then HEPES buffer was used to dilute the reaction system to 200 µl. For each reaction, 133 µl diluted reaction mixture was transferred to a 96-well plate (Greiner Bio-One), followed by the addition of 7 µl 1 mM 7-diethylamino-3-(4-maleimidophenyl)-4-methylcoumarin (CPM, Sigma). The mixture was incubated overnight at 4 °C before fluorescence detection using a SpectraMax iD5 Multi-Mode Microplate Reader (excitation, 390 nm; emission, 469 nm). Three independent experiments were conducted, and the fluorescence intensity of the acyl-CoA-free reaction system for the wild type and for each mutant were subtracted to remove the background noise. Nonlinear regression to the Michaelis–Menten equation and allosteric sigmoidal analysis were performed using GraphPad Prism 5.

LC–MS analysis of cholesterol esterification activity

The reaction system was the same as described in 'Fluorescence-based catalytic assay'. Reactions were terminated by the addition of 200 µl chloroform:methanol (2:1) and 100 ng 18:1 cholesteryl-*d*₇ ester

(Avanti) was introduced as an internal standard. Samples were vortexed and centrifuged at 2,300 g for 5 min at room temperature twice. The chloroform layer was collected, dried and resuspended in 40 µl chloroform:methanol (2:1). Serially diluted cholesteryl oleate was used as a concentration standard. Standard and samples were injected onto a 1 mm × 75 mm C18 column (ACE 3 C18, MAC-MOD) using a Shimadzu HPLC system and PAL auto-sampler. The injection volume was 20 µl and the flow rate was 70 µl min⁻¹. The column was maintained at 45 °C using a column oven. The column was connected inline to an electrospray source coupled to an LTQ-Orbitrap XL mass spectrometer (Thermo Fisher). MRFA (Met-Arg-Phe-Ala acetate salt; 2 pmol µl⁻¹ in 50% acetonitrile with 0.1% formic acid) was infused as a lock mass through a tee at the column outlet using an HPLC pump at 5 µl min⁻¹ (LC Packing). Buffer A was 0.1% formic acid and 0.028% ammonium hydroxide in acetonitrile:methanol:water (2:2:1), and buffer B was 0.1% formic acid and 0.028% ammonium hydroxide in isopropanol. Chromatographic separation was achieved with a linear gradient from 0.9% B to 100% B in 15 min and followed by a 5-min wash at 100% B and equilibrated for 10 min with 0.9% B (total programme length of 30 min). Electrospray ionization (positive mode) was achieved using a spray voltage of 4.50 kV aided by a sheath gas (nitrogen) flow rate of 18 (arbitrary units). Full scan mass spectrometry data were acquired in the Orbitrap at a resolution of 60,000 in profile mode from the *m/z* range of 385–685. Raw files were imported into Skyline software (v.4.2; University of Washington) and the peak area for lipids was extracted using the small-molecule workflow.

LC–MS analysis of lipid extractants from the enzymes

To prepare samples for LC–MS analysis, 100 µl of purified ACAT1 (2.5 mg ml⁻¹) in the W1 buffer (25 mM Tris pH 8.0, 150 mM NaCl and 0.02% GDN) was added into 800 µl extraction buffer (methanol:trichloromethane:H₂O; 4:1:3). After centrifugation at 13,000g for 10 min, the supernatant was transferred into 400 µl 100% methanol followed by 10 min centrifugation at 13,000g again. The supernatant was dried and re-dissolved in 100 µl LC buffer (methanol:water; 7:3) for further LC–MS analysis.

UHPLC (Thermo Vanquish LC system) coupled with tandem ESI-Q-Orbitrap MS (Thermo Q-Exactive Plus) was used to analyse 2 µl of lipid extractants. An Acquity UPLC BEH C₁₈ column (50 mm × 2.1 mm, 1.7 µm, Waters) was used with a gradient mobile phase of solvent A (10 mM ammonium acetate in ultra-pure water) and solvent B (methanol) at a flow rate of 0.2 ml min⁻¹. The UHPLC elution conditions were optimized as follows: 2% B (0–1 min), linear gradient from 2% to 65% B (1–12 min), 65% B (12–15 min), 65% to 100% B (15–20 min), 100% B (20–25 min), 100% to 2% B (25–28 min) and 2% B (28–30 min). The eluted lipids were directly introduced into the mass spectrometer with a vaporizer temperature of 350 °C and a spray voltage of 3,500 V under negative electrospray ionization. The analytes were monitored under full-scan data-dependent MS² mode with mass ranging from *m/z* 100 to 1,200. The MS/MS analysis was conducted with a collision energy of 35 V, a resolution of 70,000 (full width half maximum), an automatic gain control target of 5×10^4 , a maximum injection time of 200 ms and an isolation window of *m/z* 2.0. Xcalibur v.4.1.50 and Compound Discover v.3.0 were used to acquire and screen MS and MS/MS data.

Cryo-EM sample preparation and data collection

The cryo grids were prepared using Thermo Fisher Vitrobot Mark IV. Quantifoil R1.2/1.3 Cu grids were glow-discharged with air for 40 s at a medium level in Plasma Cleaner (HARRICK PLASMA, PDC-32G-2). Aliquots of 3.5 µl of purified ACAT1, concentrated to approximately 15 mg ml⁻¹, were applied to glow-discharged grids. After blotting with filter paper for 3.5 s, the grids were plunged into liquid ethane cooled with liquid nitrogen. A total of 3,921 micrograph stacks were automatically collected with SerialEM⁴¹

Article

on Titan Krios at 300 kV equipped with a K2 Summit direct electron detector (Gatan), a Quantum energy filter (Gatan) and Cs corrector (Thermo Fisher), at a nominal magnification of 105,000 \times with defocus values from $-2.0\ \mu\text{m}$ to $-1.2\ \mu\text{m}$. Each stack was exposed in super-resolution mode for 5.6 s with an exposure time of 0.175 s per frame, resulting in 32 frames per stack. The total dose rate was approximately $50\ \text{e}^- \text{\AA}^{-2}$ for each stack. The stacks were motion-corrected with MotionCor2⁴² and binned twofold, resulting in a pixel size of $1.114\ \text{\AA}$ per pixel, meanwhile dose weighting was performed⁴³. The defocus values were estimated with Gctf⁴⁴.

Cryo-EM data processing

A total of 3,084,959 particles were automatically picked with RELION^{45–47}. After 2D classification, 1,500,883 particles were selected and subject to a guided multi-reference 3D classification procedure. The references, one good and three bad, were generated with limited particles in advance. A total of 863,672 particles selected from multi-references 3D classification were subject to a global angular search 3D classification with one class and 40 iterations. The outputs of the 30th–40th iterations were applied to local angular search 3D classification with four classes separately. A total of 650,158 particles were selected by combining the good classes of the local angular search 3D classification. Next, a local search multi-reference classification procedure was performed to obtain a good class with 348,691 particles. These particles were further classified and yielded a subsets of 183,794 particles, resulting in a 3D reconstruction with an overall resolution of $3.3\ \text{\AA}$ after 3D auto-refinement with an adapted mask and C2 symmetry. The 650,158 particles were also applied for symmetry expansion using `relion_particle_symmetry_expand` in RELION⁴⁸. After further 3D classification, a total of 358,264 particles were selected to yield a dimeric reconstruction at $3.0\ \text{\AA}$ after focused refinement with an adapted mask.

All 2D classification, 3D classification and 3D auto-refinement were performed with RELION 3.0. Resolutions were estimated with the gold-standard Fourier shell correlation (FSC) 0.143 criterion⁴⁹ with high-resolution noise substitution⁵⁰. Directional FSC was also calculated for the dimeric reconstruction, which may represent a more reasonable resolution estimation⁵¹. The local resolution maps of dimeric and tetrameric ACAT1 were calculated using `relion_postprocess` in RELION.

Model building and refinement

The $3.0\ \text{\AA}$ map for the dimeric form was used to build the ACAT1 model. A previously reported crystal structure of DltB (Protein Data Bank (PDB) ID: 6BUI) was used as the initial model to be docked into the map in Chimera⁵². Manual adjustment was then performed in Coot⁵³ to generate the final structure. Two dimeric structures were fitted into the tetrameric maps to generate the tetrameric structure with a C2 symmetry. All structure refinement was carried out by PHENIX⁵⁴ in real space with secondary structure and geometry restraints. Overfitting of the dimeric model was monitored by refining the model in one of the two independent maps from the gold-standard refinement approach and testing the refined model against the other map⁵⁵. All structure figures were prepared with PyMOL⁵⁶. Structural conservation was analysed using the ConSurf server⁵⁷, on the basis of sequence alignment of ACAT1, ACAT2, DGAT1, GOAT, Porcupine and HHAT from human and mouse using Clustal Omega⁵⁸.

Statistics and reproducibility

No statistical methods were used to predetermine sample size. The experiments were not randomized, and the investigators were not blinded to allocation during experiments and outcome assessment. Each experiment was conducted independently at least twice with similar results.

Reporting summary

Further information on research design is available in the Nature Research Reporting Summary linked to this paper.

Data availability

The atomic coordinates of the tetrameric and dimeric ACAT1 have been deposited in the PDB under accession codes 6P2P and 6P2J, respectively. The corresponding electron microscopy maps have been deposited in the Electron Microscopy Data Bank under accession codes EMD-20239 and EMD-20238, respectively. For uncropped SDS–PAGE gels, see Supplementary Fig. 1. Source Data for Figs. 1–3 and Extended Data Figs. 2, 7 are provided with the paper. The raw electron micrographs for structural analysis are available from the corresponding authors upon reasonable request.

41. Mastronarde, D. N. Automated electron microscope tomography using robust prediction of specimen movements. *J. Struct. Biol.* **152**, 36–51 (2005).
42. Zheng, S. Q. et al. MotionCor2: anisotropic correction of beam-induced motion for improved cryo-electron microscopy. *Nat. Methods* **14**, 331–332 (2017).
43. Grant, T. & Grigorieff, N. Measuring the optimal exposure for single particle cryo-EM using a 2.6 \AA reconstruction of rotavirus VP6. *eLife* **4**, e06980 (2015).
44. Zhang, K. Gctf: Real-time CTF determination and correction. *J. Struct. Biol.* **193**, 1–12 (2016).
45. Scheres, S. H. Semi-automated selection of cryo-EM particles in RELION-1.3. *J. Struct. Biol.* **189**, 114–122 (2015).
46. Scheres, S. H. RELION: implementation of a Bayesian approach to cryo-EM structure determination. *J. Struct. Biol.* **180**, 519–530 (2012).
47. Kimanius, D., Forsberg, B. O., Scheres, S. H. & Lindahl, E. Accelerated cryo-EM structure determination with parallelisation using GPUs in RELION-2. *eLife* **5**, e18722 (2016).
48. Scheres, S. H. Processing of structurally heterogeneous cryo-EM data in RELION. *Methods Enzymol.* **579**, 125–157 (2016).
49. Rosenthal, P. B. & Henderson, R. Optimal determination of particle orientation, absolute hand, and contrast loss in single-particle electron cryomicroscopy. *J. Mol. Biol.* **333**, 721–745 (2003).
50. Chen, S. et al. High-resolution noise substitution to measure overfitting and validate resolution in 3D structure determination by single particle electron cryomicroscopy. *Ultramicroscopy* **135**, 24–35 (2013).
51. Dang, S. et al. Cryo-EM structures of the TMEM16A calcium-activated chloride channel. *Nature* **552**, 426–429 (2017).
52. Pettersen, E. F. et al. UCSF Chimera—a visualization system for exploratory research and analysis. *J. Comput. Chem.* **25**, 1605–1612 (2004).
53. Emsley, P., Lohkamp, B., Scott, W. G. & Cowtan, K. Features and development of Coot. *Acta Crystallogr. D* **66**, 486–501 (2010).
54. Adams, P. D. et al. PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr. D* **66**, 213–221 (2010).
55. Amunts, A. et al. Structure of the yeast mitochondrial large ribosomal subunit. *Science* **343**, 1485–1489 (2014).
56. DeLano, W. L. The PyMOL Molecular Graphics System, <http://www.pymol.org> (2002).
57. Ashkenazy, H. et al. ConSurf 2016: an improved methodology to estimate and visualize evolutionary conservation in macromolecules. *Nucleic Acids Res.* **44** (W1), W344–W350 (2016).
58. Sievers, F. et al. Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. *Mol. Syst. Biol.* **7**, 539 (2011).

Acknowledgements We thank P. Shao for technical support during electron microscopy image acquisition; S. Kyn for technical support during the mass spectrometry analysis of cholesteryl oleate; and S. Dang for assistance with directional FSC analysis. We acknowledge the use of Princeton's Imaging and Analysis Center, which is partially supported by the Princeton Center for Complex Materials, and the National Science Foundation (NSF)-MRSEC programme (DMR-1420541). This work was supported in part by the Ara Parseghian Medical Research Foundation (N.Y.). H.Q. is supported by the New Jersey Council for Cancer Research. N.Y. is supported by the Shirley M. Tilghman endowed professorship from Princeton University.

Author contributions N.Y., R.Y. and H.Q. conceived and N.Y. supervised the project. H.Q. and X.Z. designed the experiments. H.Q., X.Z. and R.Y. performed cloning and protein purification. H.Q. prepared cryo-EM samples, collected data and determined the structures. X.Z. and H.Q. performed the fluorescence-based activity assays. X.Z. validated the fluorescence-based assay by detecting cholesterol ester by mass spectrometry. X.S. and C.C.L.W. analysed lipid extractants from the enzymes by LC–MS. S.G. and X.Y. performed molecular docking of cholesterol. X.D. and H.Y. contributed to data analysis. N.Y., H.Q. and X.Z. wrote the manuscript.

Competing interests The authors declare no competing interests.

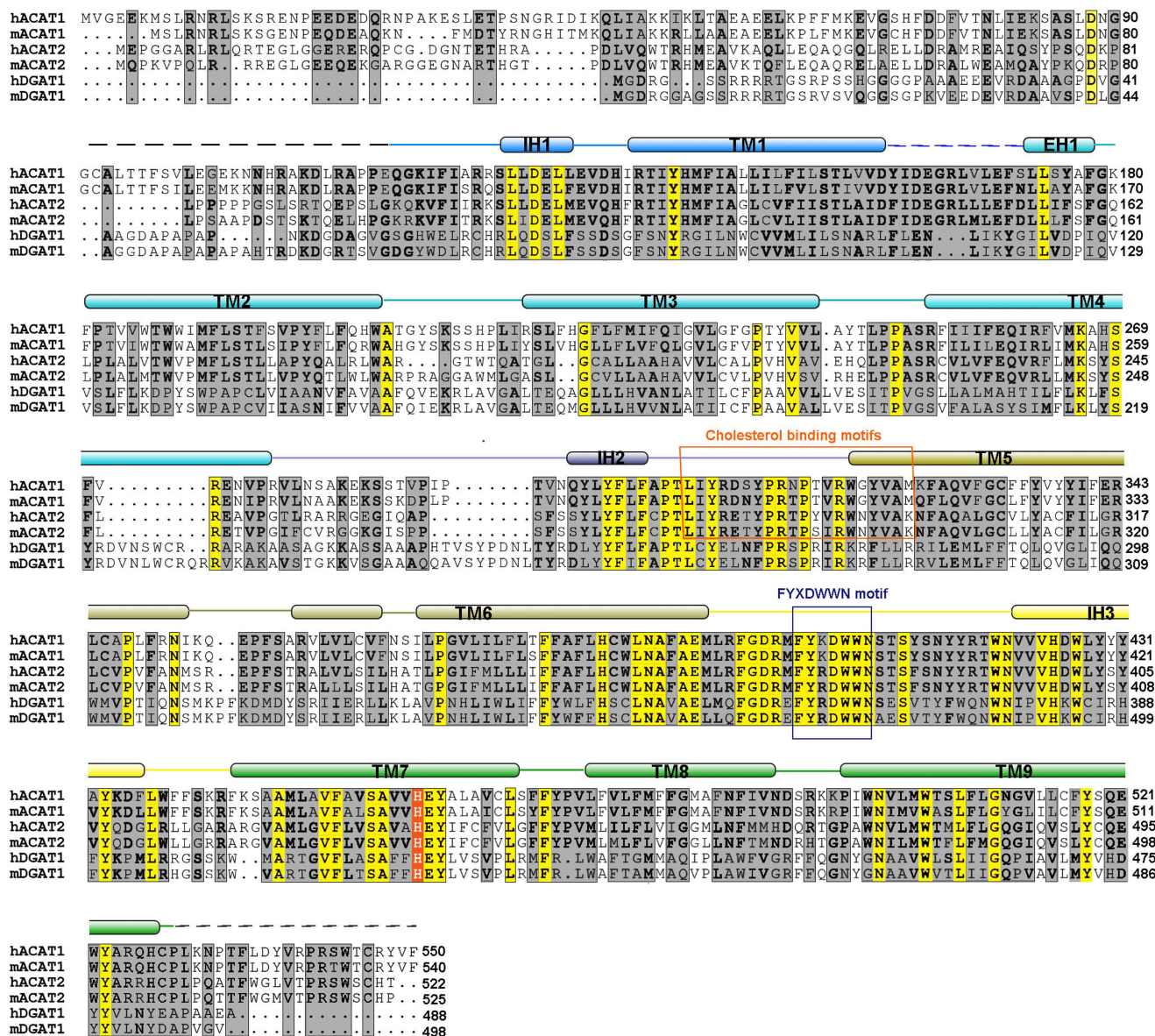
Additional information

Supplementary information is available for this paper at <https://doi.org/10.1038/s41586-020-2290-0>.

Correspondence and requests for materials should be addressed to H.Q. or N.Y.

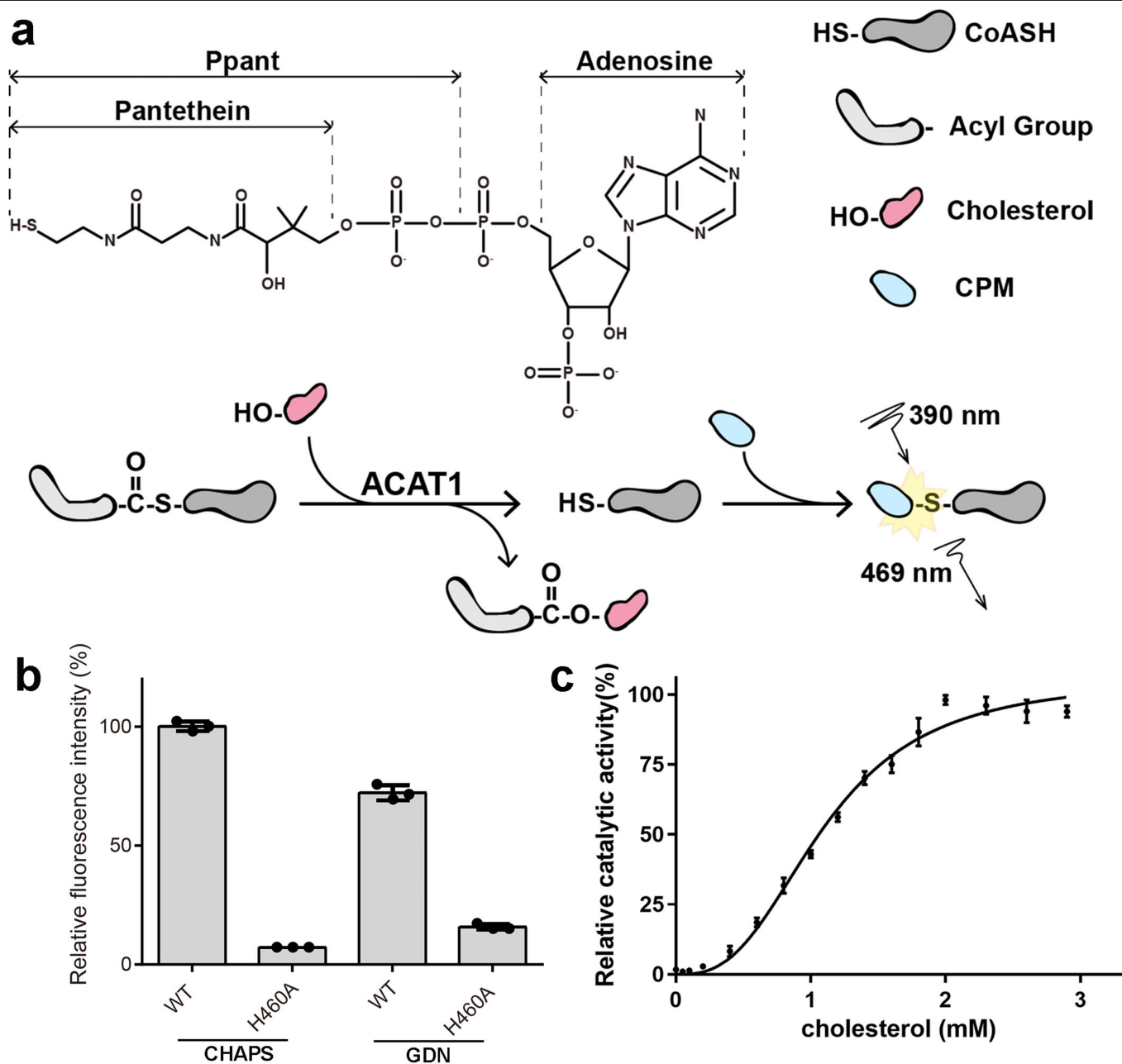
Peer review information Nature thanks David Drew, Savvas N. Savvides and the other, anonymous, reviewer(s) for their contribution to the peer review of this work.

Reprints and permissions information is available at <http://www.nature.com/reprints>.



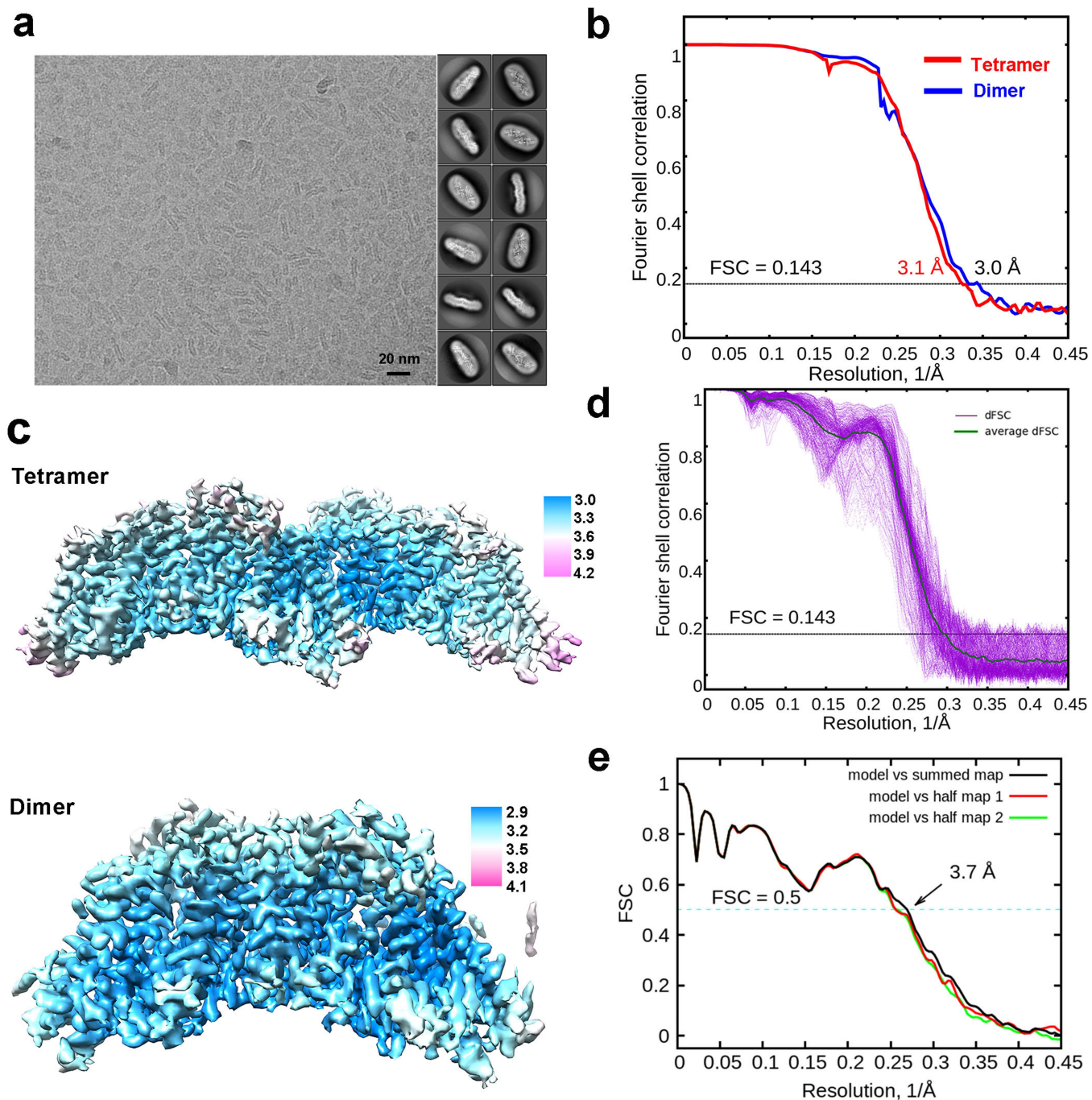
Extended Data Fig.1 | Sequence alignment of human and mouse ACAT1, ACAT2 and DGAT1. Secondary structural elements of human ACAT1 are indicated above the sequences according to the present cryo-EM structure. Invariant and highly conserved residues are shaded yellow and grey, respectively. The conserved residue His460 at the active site is shaded red and

coloured white. The conserved FYXDWWN motif and predicted cholesterol binding motif are indicated with blue and orange boxes, respectively. Sequences from human (h) and mouse (m) were aligned using the online MultiAlin server (<http://multalin.toulouse.inra.fr>).



Extended Data Fig. 2 | Enzymatic activity of recombinant human ACAT1.
a, Schematic of our fluorescence-based activity assay for ACAT1. Top, the chemical structure of CoASH; bottom, a schematic illustration of the fluorescence-based activity assay. **b**, Interference of detergents on the enzymatic activity of ACAT1. The proteins used for the assay were purified by

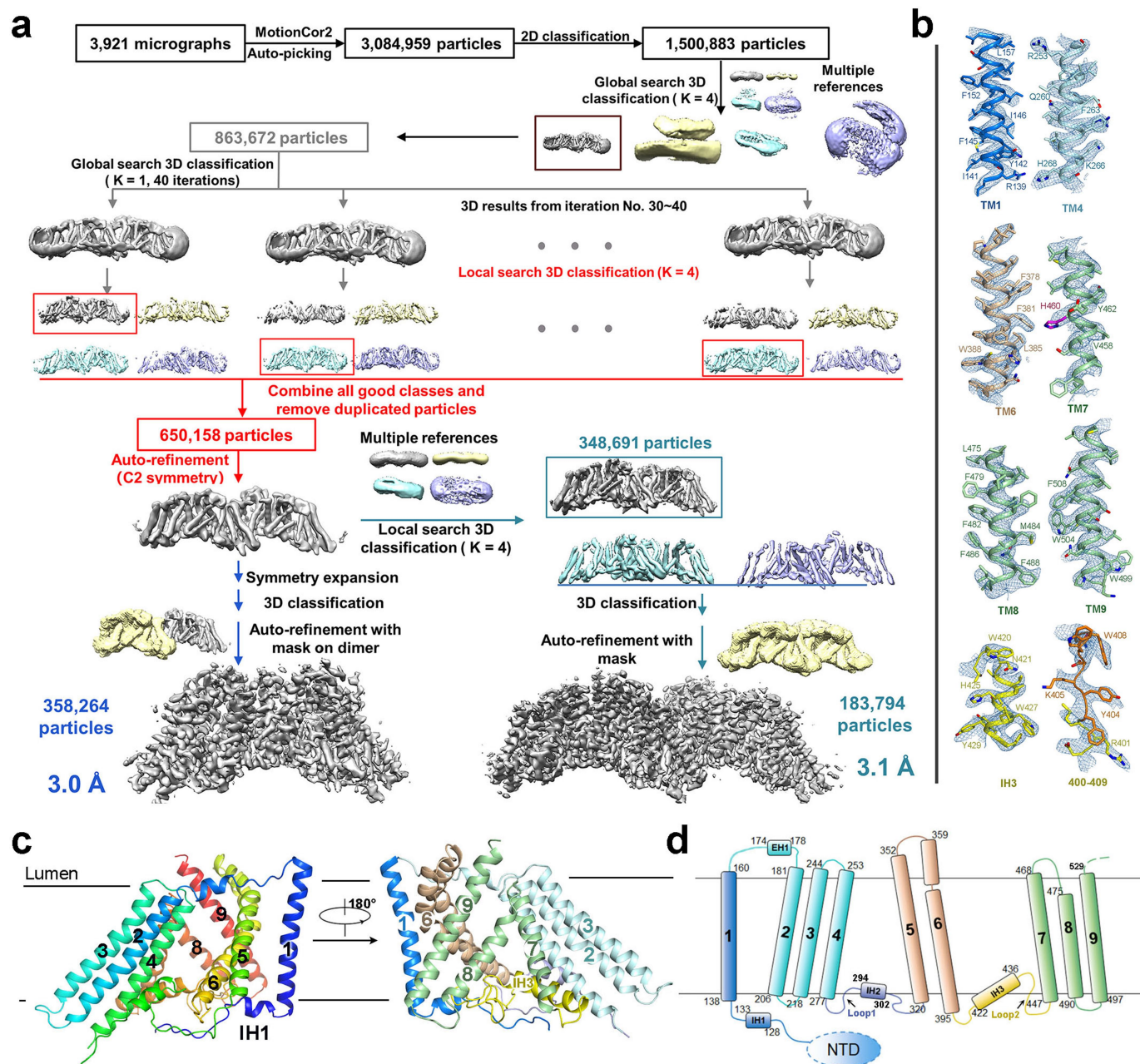
SEC in the presence of 1% CHAPS or 0.02% GDN. **c**, Allosteric activation of ACAT1 by cholesterol. The sigmoidal plot of catalytic activity with increasing concentrations of cholesterol is consistent with the proposed allosteric activation of ACAT1 by cholesterol³². Data in **b**, **c** are mean \pm s.d. of three independent experiments.



Extended Data Fig. 3 | Cryo-EM analysis of the structure of human ACAT1.

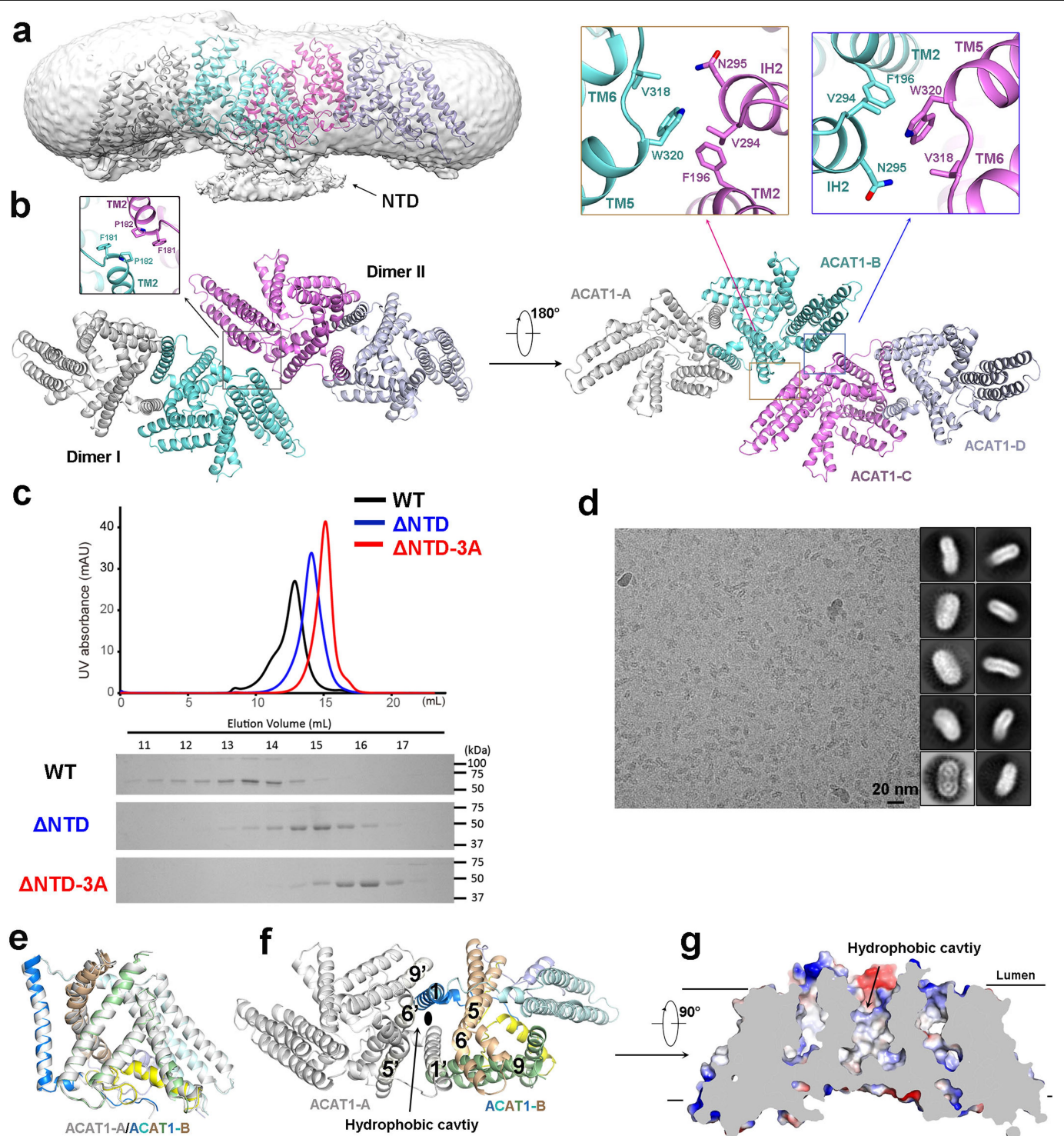
a, A representative micrograph (left) and 2D class averages (right) of cryo-samples of ACAT1 in GDN micelles. The box size for 2D averages is 310 Å. **b**, FSC curves for the 3D electron microscopy reconstructions of tetrameric and dimeric ACAT1. **c**, Local resolution map of tetrameric (top) and dimeric (bottom) ACAT1 calculated using RELION 3.0. The resolution bars on the right are labelled in Å. **d**, Directional FSC (dFSC) for the dimeric reconstruction. Each

purple curve indicates a different direction. In total, 500 dFSC curves were generated, which were averaged and shown by the green curve (average dFSC)⁵¹. **e**, FSC curves of the refined model versus the summed map that it was refined against (black); of the model refined in the first of the two independent maps used for the gold-standard FSC versus that same map (red); and of the model refined in the first of the two independent maps versus the second independent map (green) for the dimeric reconstruction.



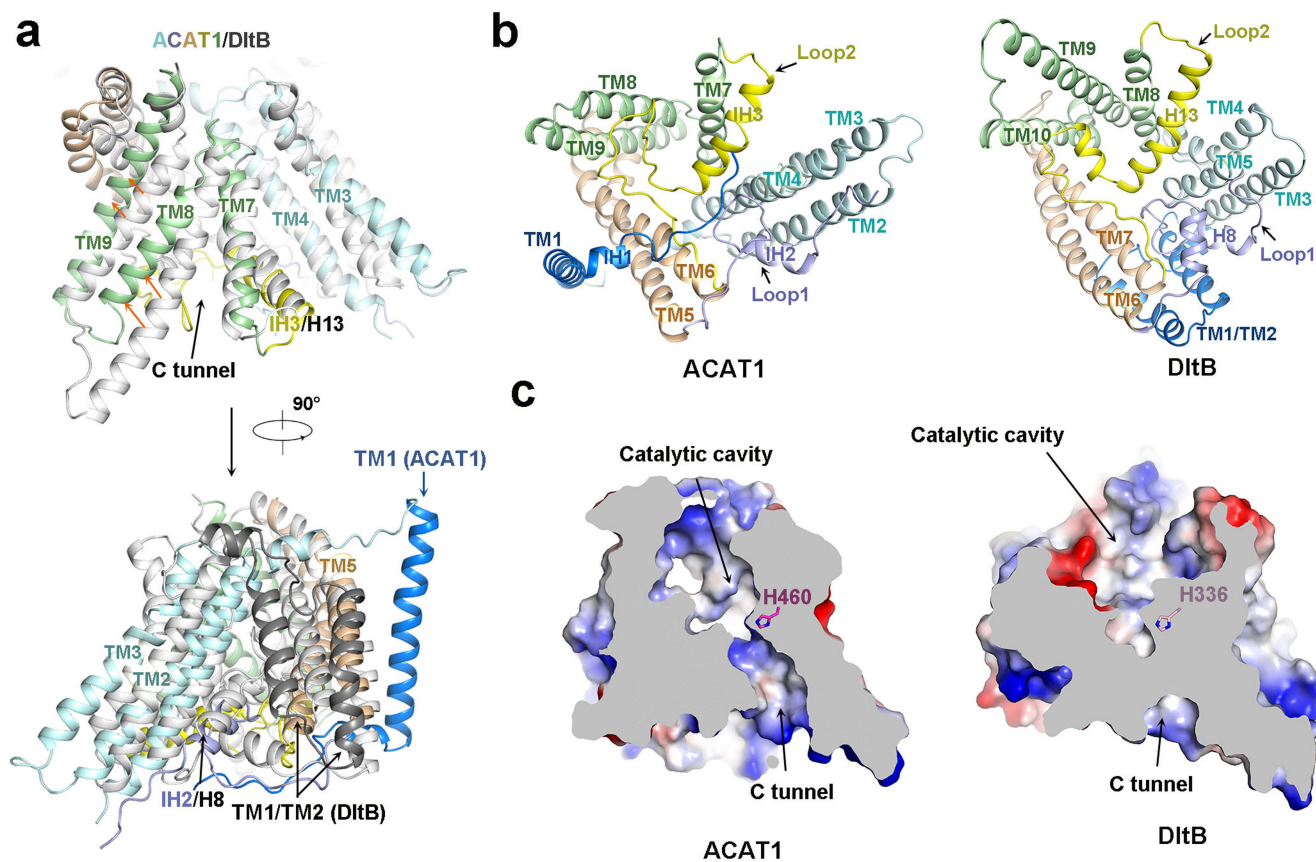
Extended Data Fig. 4 | Flowchart for structural determination. a, Flowchart of data processing; see Methods for details. **b**, Electron microscopy maps of representative structural elements. The densities, contoured at 10–13 σ , were prepared in PyMOL. **c**, Structure of an ACAT1 protomer. The structure is

rainbow-coloured on the left (blue for the amino terminus and red for the carboxyl terminus) and domain-coloured on the right. **d**, Topological structure of ACAT1. The structural elements are colour-coded to match the domain colours in **c**.



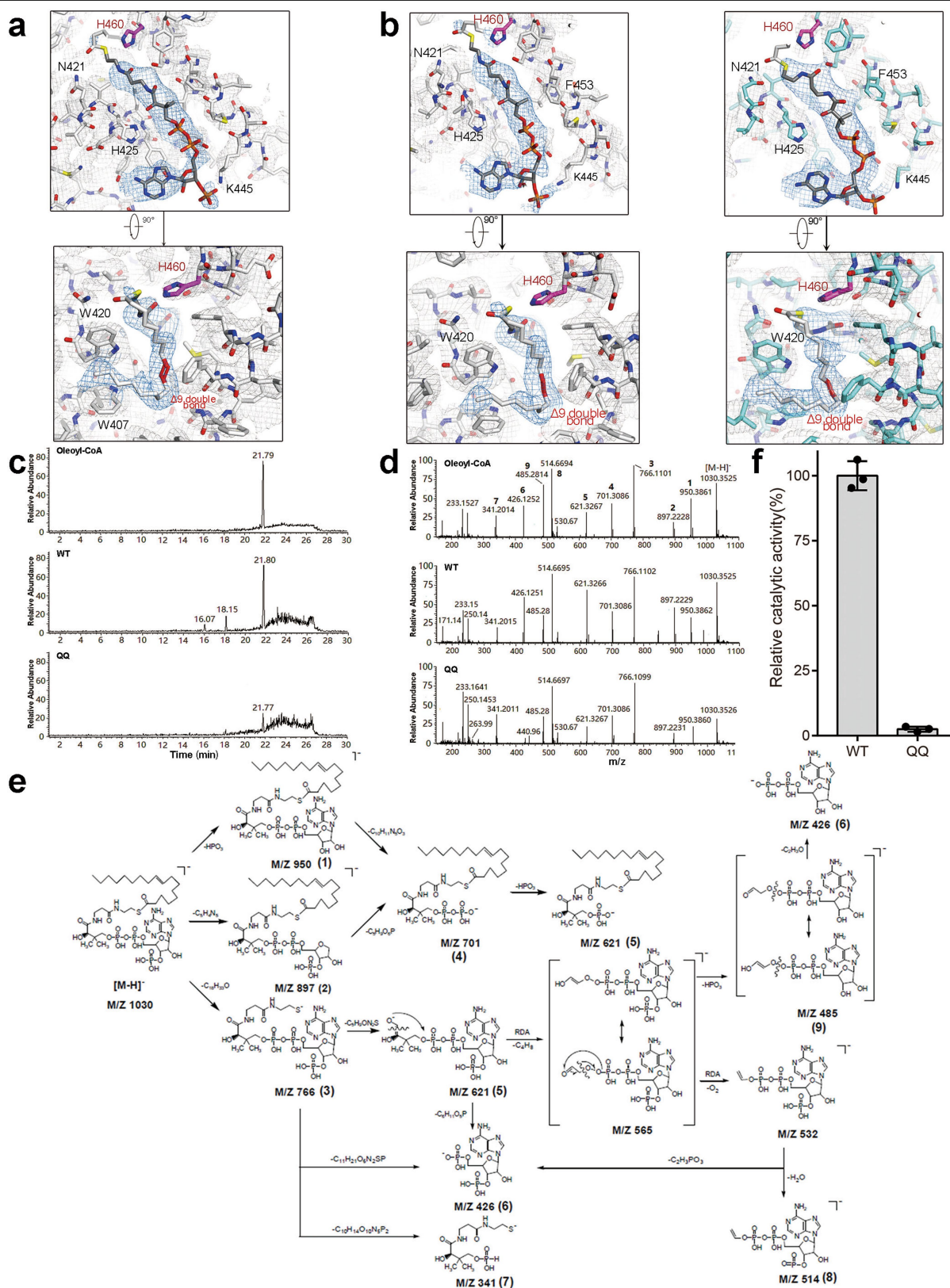
Extended Data Fig. 5 | NTD is responsible for tetramerization. **a**, Electron microscopy map of the tetrameric ACAT1, displayed at low threshold (0.004) in Chimera, reveals extra cytosolic densities that may belong to the NTD. **b**, Tetrameric ACAT1 shown in the luminal (left) and cytosolic (right) views. The insets show residues on the tetrameric interface. **c**, Validation of the oligomeric states of dimeric and monomeric mutants using SEC. SEC profiles and corresponding SDS-PAGE gels for wild-type ACAT1 and two variants, ACAT1(ΔNTD) and ACAT1(ΔNTD-3A), in GDN micelles are shown. The experiment was independently repeated twice with similar results.

d, A representative micrograph (left) and representative 2D averages (right) of ACAT1(ΔNTD). The box size for the 2D averages is 220 Å, whereas that for wild-type ACAT1 is 310 Å. **e**, The two protomers in each dimer are nearly identical. Superimposition of the two protomers in one dimer is shown. **f**, Luminal view of the dimeric ACAT. An open cavity is formed by TM1, TM5, TM6 and TM9 from two protomers around the C2 axis, which is indicated by the black oval in the centre. **g**, The luminal cavity in the centre of each dimer is highly hydrophobic. The electrostatic surface potential, calculated in PyMOL, is shown in a cut-open side view.



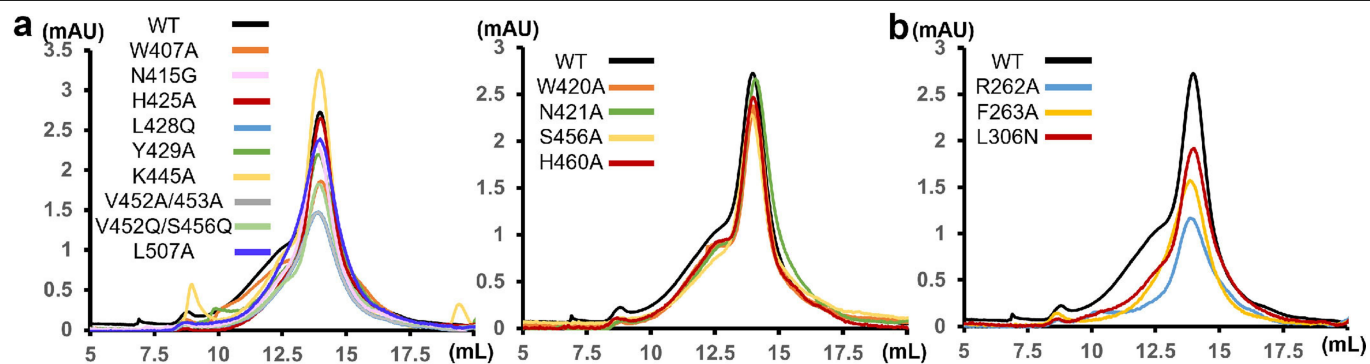
Extended Data Fig. 6 | Structural comparison of ACAT1 and DltB. **a**, ACAT1 and DltB share an identical structural core. TM2–TM9 of ACAT1 can be superimposed onto TM3–TM10 of DltB (PDB ID: 6BUI) with an r.m.s.d. of 5.0 Å over 272 C α atoms. Superimposition of ACAT1 onto DltB is shown in two perpendicular side views. The major conformational shifts of TM8 and TM9 in ACAT1 from the corresponding segments in DltB are indicated with orange arrows. TM1 of ACAT1 and the corresponding segments TM1 and TM2 (dark

grey) in DltB adopt different structures. **b**, Loop1 and Loop2 constitute the major cytosolic segments in both ACAT1 and DltB. The cytosolic views of the two proteins, with corresponding structural segments coloured the same, are shown here. **c**, There is no C tunnel in DltB as there is in ACAT1. The electrostatic surface potentials of ACAT1 and DltB are shown in the same cut-open side views. The conserved His residue is shown as magenta sticks in both structures.

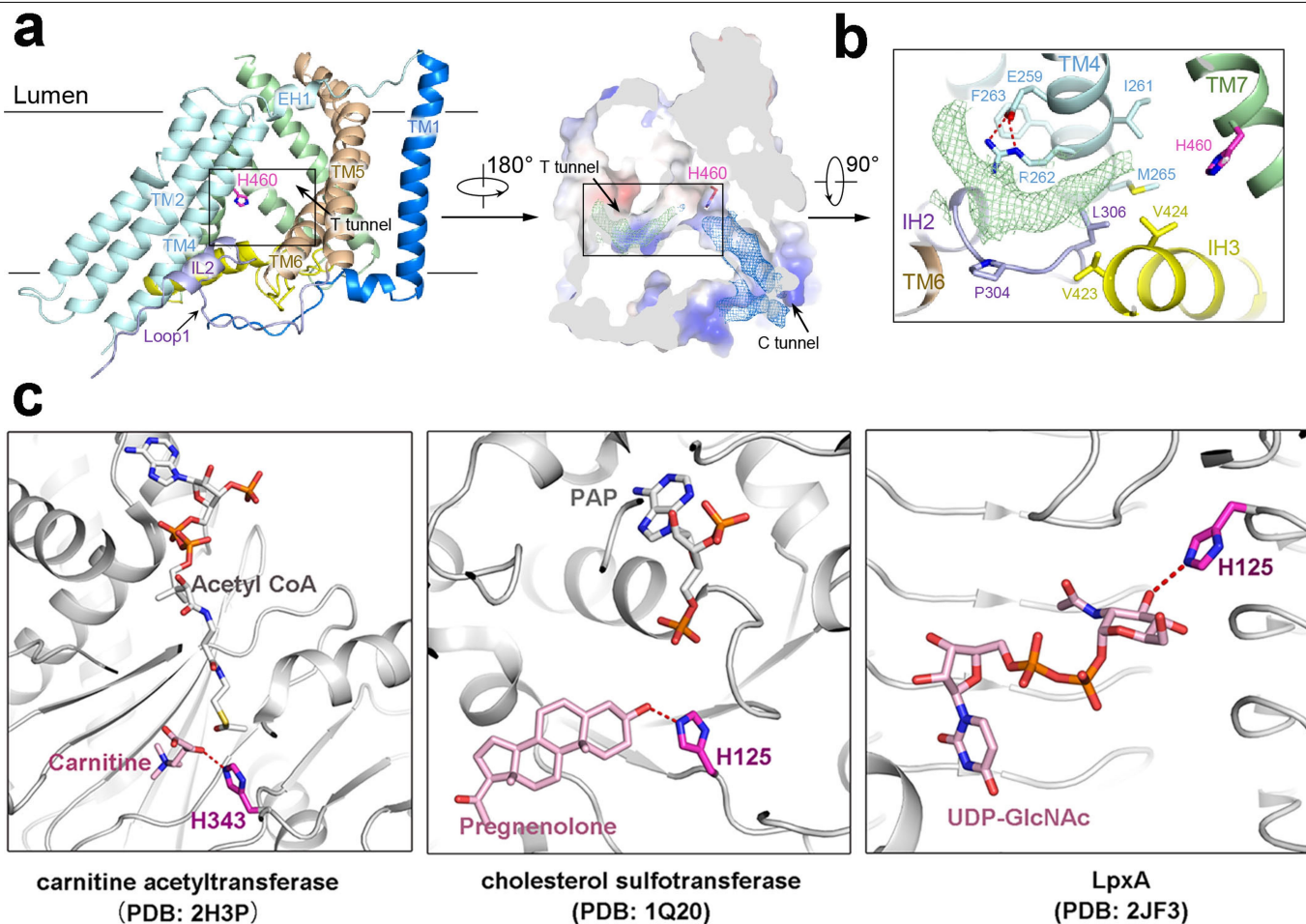


Extended Data Fig. 7 | LC-MS identification of the ligand to which the linear density in the structure might belong. a, Electron microscopy densities for oleoyl-CoA in the ACAT1-A protomer from the dimer reconstruction. The densities for oleoyl-CoA (shown as blue mesh) and surrounding residues (shown as grey mesh) are contoured at 6σ . Two perpendicular views are shown. **b**, Electron microscopy densities for oleoyl-CoA in the tetrameric reconstruction. All the densities were contoured at 5σ . Two perpendicular views of ACAT1-A (left) and ACAT1-B (right) are shown. The densities in the

other two protomers are not shown because of the C2 symmetry. **c**, LC profiles of commercial oleoyl-CoA (top), and lipids extracted from wild-type enzymes (middle) and the QQ mutant (bottom). **d**, MS/MS spectrum of commercial oleoyl-CoA (top) and extracted oleoyl-CoA from wild-type enzymes (middle) and QQ mutant (bottom). Fractions 1-10 represent the same fragments as those in **e**. **e**, Potential MS/MS fragmentation pattern of oleoyl-CoA. **f**, The QQ mutant shows nearly complete loss of enzymatic activity. Data are mean \pm s.d. of three independent experiments.



Extended Data Fig. 8 | SEC profiles of the ACAT1 mutants in activity assays. a, SEC profiles of enzymes with mutations related to oleoyl-CoA coordination. **b,** SEC profiles of enzymes with mutations related to the T tunnel. The experiments were independently repeated twice with similar results.



Extended Data Fig. 9 | The T tunnel may serve as the cholesterol entry site.
a, Cholesterol may access the active site through the T tunnel. Left, a side view of one protomer looking through the T tunnel. The black box indicates the position of the T tunnel. Right, a stretched density is found in the T tunnel. The contour of the density is not reminiscent of cholesterol or GDN. It may result from a mixture of molecules. Nevertheless, the presence of such density suggests that a hydrophobic molecule can enter this tunnel. The density, shown as green mesh, is contoured at 6σ . The density for the potentially bound oleoyl-CoA (blue mesh) is also shown at 6σ as a reference. **b**, Residues

constituting the T tunnel. The density is shown to indicate the tunnel. **c**, A conserved histidine residue is found in the active site in the crystal structures of carnitine acetyltransferase (PDB ID: 2H3P), cholesterol sulfotransferase (PDB ID: 1Q20), and UDP-*N*-acetylglucosamine acyltransferase (LpxA) (PDB ID: 2JF3). This residue is highlighted as magenta sticks in all three panels. The bound substrates—carnitine, pregnenolone and UDP-GlcNAc—are all coloured light pink. The crucial histidine residue may activate the nucleophilic substrate through deprotonation.

Extended Data Table 1 | Data collection, 3D reconstruction and model statistics

	Dimeric ACAT1 (EMDB-20238) (PDB 6P2J)	Tetrameric ACAT1 (EMDB-20239) (PDB 6P2P)
Data collection and processing		
Magnification	105,000	105,000
Voltage (kV)	300	300
Electron exposure (e-/Å ²)	50	50
Defocus range (µm)	-2.0~-1.2	-2.0~-1.2
Pixel size (Å)	1.114	1.114
Symmetry imposed	C1	C2
Initial particle images (no.)	3,084,959	3,084,959
Final particle images (no.)	358,264	183,794
Map resolution (Å)	3.0	3.1
FSC threshold	0.143	0.143
Map resolution range (Å)	2.9~4.1	3.0~4.2
Refinement		
Initial model used (PDB code)	6BUI	6P2J
Model resolution (Å)	3.7	3.8
FSC threshold	0.5	0.5
Model resolution range (Å)	---	---
Map sharpening <i>B</i> factor (Å ²)	-92	-86
Model composition		
Non-hydrogen atoms	7015	13966
Protein residues	816	1632
Ligands	2	4
<i>B</i> factors (Å ²)		
Protein	86.11	87.00
Ligand	101.67	102.34
R.m.s. deviations		
Bond lengths (Å)	0.002	0.005
Bond angles (°)	1.98	0.80
Validation		
MolProbity score	1.77	3.07
Clashscore	3.42	17.07
Poor rotamers (%)	1.78	0.63
Ramachandran plot		
Favored (%)	93.07	91.83
Allowed (%)	6.06	7.92
Disallowed (%)	0.87	0.25

Reporting Summary

Nature Research wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Research policies, see [Authors & Referees](#) and the [Editorial Policy Checklist](#).

Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

n/a Confirmed

- ☐ ☒ The exact sample size (n) for each experimental group/condition, given as a discrete number and unit of measurement
- ☐ ☒ A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
- ☒ ☐ The statistical test(s) used AND whether they are one- or two-sided
Only common tests should be described solely by name; describe more complex techniques in the Methods section.
- ☒ ☐ A description of all covariates tested
- ☒ ☐ A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
- ☐ ☒ A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)
- ☒ ☐ For null hypothesis testing, the test statistic (e.g. F , t , r) with confidence intervals, effect sizes, degrees of freedom and P value noted
Give P values as exact values whenever suitable.
- ☒ ☐ For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
- ☒ ☐ For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
- ☒ ☐ Estimates of effect sizes (e.g. Cohen's d , Pearson's r), indicating how they were calculated

Our web collection on [statistics for biologists](#) contains articles on many of the points above.

Software and code

Policy information about [availability of computer code](#)

Data collection

SerialEM 3.7.2, Xcalibur 4.1.50

Data analysis

RELION 3.0, MotionCor2 1.0.5, GCTF 1.06, Phenix 1.16, Coot 0.8.9, Pymol 4.6.0, Chimera 1.12, GraphPad Prism 7.00, SnapGene 4.2.6, Skyline 4.2, Compound Discover 3.0, Schrödinger Suite 2017-1

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors/reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Research [guidelines for submitting code & software](#) for further information.

Data

Policy information about [availability of data](#)

All manuscripts must include a [data availability statement](#). This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A list of figures that have associated raw data
- A description of any restrictions on data availability

The atomic coordinates of the dimeric and tetrameric model have been deposited in the Protein Data Bank (<http://www.rcsb.org>) under the accession code 6P2J and 6P2P, respectively. The corresponding EM maps have been deposited in the Electron Microscopy Data Bank (<https://www.ebi.ac.uk/pdbe/emdb/>) with the accession code EMDB-20238, and EMDB-20239, respectively. All other data are available from the corresponding author upon reasonable request. Correspondence and requests for materials should be addressed to N.Y. (nyan@princeton.edu) and H.Q. (hongwuq@princeton.edu).

Field-specific reporting

Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

☒ Life sciences ☐ Behavioural & social sciences ☐ Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see [nature.com/documents/nr-reporting-summary-flat.pdf](https://www.nature.com/documents/nr-reporting-summary-flat.pdf)

Life sciences study design

All studies must disclose on these points even when the disclosure is negative.

Sample size	For all the functional assays, three independent experiments (n=3) were carried out.
Data exclusions	None.
Replication	For the biochemical assay, each experiment was repeated 3 times. All data could be reproduced.
Randomization	n/a. The samples were not randomized for the experiments.
Blinding	n/a. No blinding used in this study.

Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

Materials & experimental systems

n/a	Involved in the study
<input type="checkbox"/>	<input checked="" type="checkbox"/> Antibodies
<input type="checkbox"/>	<input checked="" type="checkbox"/> Eukaryotic cell lines
<input checked="" type="checkbox"/>	<input type="checkbox"/> Palaeontology
<input checked="" type="checkbox"/>	<input type="checkbox"/> Animals and other organisms
<input checked="" type="checkbox"/>	<input type="checkbox"/> Human research participants
<input checked="" type="checkbox"/>	<input type="checkbox"/> Clinical data

Methods

n/a	Involved in the study
<input checked="" type="checkbox"/>	<input type="checkbox"/> ChIP-seq
<input checked="" type="checkbox"/>	<input type="checkbox"/> Flow cytometry
<input checked="" type="checkbox"/>	<input type="checkbox"/> MRI-based neuroimaging

Antibodies

Antibodies used	anti-FLAG M2 affinity resin (Sigma-A2220-SLBX3759)
Validation	https://www.sigmaaldrich.com/catalog/CertOfAnalysisPage.do?symbol=A2220&LotNo=SLBX3759&brandTest=SIGMA&returnUrl=%2Fproduct%2FSIGMA%2FA2220

Eukaryotic cell lines

Policy information about [cell lines](#)

Cell line source(s)	HEK293F (invitrogen)
Authentication	No further authentication was performed for commercially available cell lines.
Mycoplasma contamination	Not tested for mycoplasma contamination.
Commonly misidentified lines (See ICLAC register)	No commonly misidentified cell lines were used.

Structure of nevanimibe-bound tetrameric human ACAT1

<https://doi.org/10.1038/s41586-020-2295-8>

Tao Long^{1,3}, Yingyuan Sun^{1,3}, Abdirahman Hassan¹, Xiaofeng Qi¹ & Xiaochun Li^{1,2}✉

Received: 26 January 2020

Accepted: 17 March 2020

Published online: 13 May 2020

 Check for updates

Cholesterol is an essential component of mammalian cell membranes, constituting up to 50% of plasma membrane lipids. By contrast, it accounts for only 5% of lipids in the endoplasmic reticulum (ER)¹. The ER enzyme sterol *O*-acyltransferase 1 (also named acyl-coenzyme A:cholesterol acyltransferase, ACAT1) transfers a long-chain fatty acid to cholesterol to form cholesteryl esters that coalesce into cytosolic lipid droplets. Under conditions of cholesterol overload, ACAT1 maintains the low cholesterol concentration of the ER and thereby has an essential role in cholesterol homeostasis^{2,3}. ACAT1 has also been implicated in Alzheimer's disease⁴, atherosclerosis⁵ and cancers⁶. Here we report a cryo-electron microscopy structure of human ACAT1 in complex with nevanimibe⁷, an inhibitor that is in clinical trials for the treatment of congenital adrenal hyperplasia. The ACAT1 holoenzyme is a tetramer that consists of two homodimers. Each monomer contains nine transmembrane helices (TMs), six of which (TM4–TM9) form a cavity that accommodates nevanimibe and an endogenous acyl-coenzyme A. This cavity also contains a histidine that has previously been identified as essential for catalytic activity⁸. Our structural data and biochemical analyses provide a physical model to explain the process of cholesterol esterification, as well as details of the interaction between nevanimibe and ACAT1, which may help to accelerate the development of ACAT1 inhibitors to treat related diseases.

Cholesterol is an essential structural component of cell membranes as well as a precursor for bile acids and steroid hormones⁹. Cells obtain cholesterol from endogenous biosynthesis in the ER¹⁰ and from exogenous delivery via low-density-lipoprotein (LDL) receptors³. The ER enzyme ACAT1 transfers a fatty acyl group from acyl-coenzyme A (acyl-CoA) to the 3 β -hydroxyl moiety of cholesterol² (Fig. 1a); the resulting cholesteryl esters then coalesce to form cytoplasmic lipid droplets that store and sequester cholesterol, preventing the formation of cholesterol crystals that are lethal to cells¹¹.

ACAT1 belongs to the membrane-bound *O*-acyltransferase (MBOAT) enzyme family, the members of which are distributed widely in prokaryotic and eukaryotic cells^{12,13}. Eukaryotic MBOATs—such as Hedgehog acyltransferase, ghrelin acyltransferase and Porcupine—attach fatty acids to proteins or small molecules^{12,13}. Although all MBOATs catalyse acyl transfer, they exhibit very little sequence identity. A recent study has detailed the structure of an MBOAT that modifies protective cell-wall polymers in bacteria¹⁴; however, to our knowledge there have been no structures of eukaryotic MBOATs reported to date.

There are two ACATs in humans^{15–18}, designated ACAT1 and ACAT2 (Extended Data Fig. 1). ACAT1 is found in all nucleated eukaryotic cells, where its products are incorporated into cytosolic lipid droplets. ACAT2 is produced primarily in hepatocytes and intestinal epithelial cells, and its products are translocated to the ER lumen where they are incorporated into secreted lipoproteins. When purified in detergents, full-length ACAT1 forms a homotetramer, whereas truncated

ACAT1—with a cytosolic N-terminal 65-amino-acid deletion—forms a homodimer¹⁹. Notably, the maximum rate of reaction (V_{\max}) of the homodimer is fivefold higher than that of the homotetramer¹⁹. Mutagenesis studies have identified His460, an intramembrane histidine residue located in what has been putatively assigned as TM7, as a requirement for catalytic activity⁸. This finding led to a proposed mechanism in which His460 mediates the transfer of the fatty acyl group. Previous studies have shown that blocking ACAT1 alleviates neurological damage in a mouse model of Alzheimer's disease⁴. In addition, reducing ACAT1 activity was found to decrease plaque size and cholesteryl ester content in the aortas of hypercholesteremic mice⁵; by contrast, a clinical study showed that the ACAT inhibitor pactimibe could not slow atherosclerosis in patients with coronary disease²⁰. Pharmacological inhibition of ACAT1 has been shown to reduce the proliferation of cancer cells in vitro⁶. These findings demonstrate the potential of ACAT1 as a therapeutic target, and a detailed knowledge of its structure should aid in the search for more potent inhibitors.

Structure of tetrameric ACAT1

We expressed full-length human ACAT1 in HEK293 cells and measured its activity in vitro using a previously reported fluorescence-based acyltransferase assay²¹. In brief, we incubated recombinant ACAT1 with oleoyl-CoA and mixed micelles containing 1-palmitoyl-2-oleoyl-*sn*-glycero-3-phosphocholine (POPC) with or without

¹Department of Molecular Genetics, University of Texas Southwestern Medical Center, Dallas, TX, USA. ²Department of Biophysics, University of Texas Southwestern Medical Center, Dallas, TX, USA.

³These authors contributed equally: Tao Long, Yingyuan Sun. ✉e-mail: xiaochun.li@utsouthwestern.edu

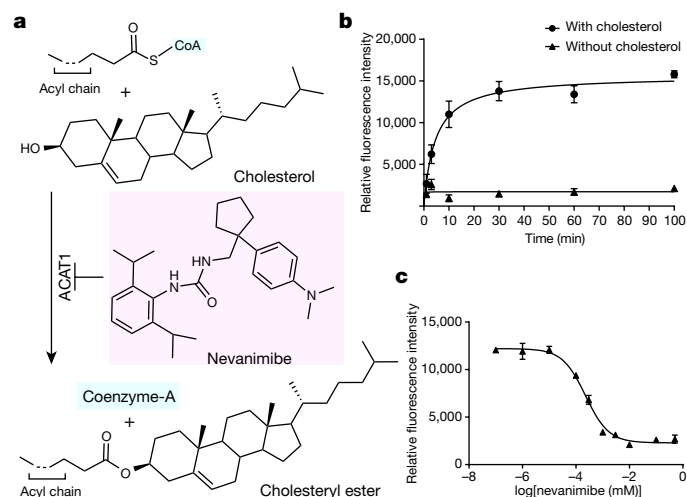


Fig. 1 | Functional characterization of human ACAT1 and nevanimibe in vitro.

a, The cholesterol esterification reaction. The chemical structures of cholesterol, acyl-CoA and nevanimibe are shown. **b**, Time curve of cholesterol esterification by ACAT1 with oleoyl-CoA. No activity was detected when cholesterol was absent from the mixed micelles. **c**, Nevanimibe inhibits the activity of ACAT1 in vitro, with an IC_{50} value of $0.23 \pm 0.06 \mu M$ ($\log IC_{50} = -3.641 \pm 0.123$). ACAT1 activity was measured by monitoring the released sulfhydryl (-SH) group of CoA. The fluorescent product was detected with excitation and emission wavelengths of 355 nm and 460 nm, respectively. Data are mean \pm s.d. ($n = 3$ technically independent experiments). Each experiment was reproduced at least twice on separate occasions with similar results.

cholesterol. After incubation at 37 °C we added 7-diethylamino-3-(4-maleimidophenyl)-4-methylcoumarin (CPM), which reacts with the sulfhydryl group of free CoA to form a fluorescent adduct. Recombinant ACAT1 demonstrated the expected biological activity in the presence

of cholesterol in vitro (Fig. 1b), and nevanimibe was found to inhibit the cholesterol esterification reaction with a half-maximal inhibitory concentration (IC_{50}) of approximately $0.23 \mu M$ (Fig. 1a, c).

We co-purified recombinant ACAT1 with nevanimibe in detergent, without the addition of cholesterol, acyl-CoA or any other fatty acid. The complex eluted as a single peak in the size-exclusion chromatography profile, rendering it suitable for structural investigation (Extended Data Fig. 2). We then determined the structure of ACAT1 at 3.67 Å resolution by cryo-electron microscopy (cryo-EM) (Fig. 2a, Extended Data Figs. 3–6, Extended Data Table 1). The ACAT1 holoenzyme consists of two dimers that are related by C2 symmetry (Fig. 2b); we therefore assumed C2 symmetry during data processing. We termed the four ACAT1 molecules ACAT1-A and ACAT1-B in dimer 1, and ACAT1-C and ACAT1-D in dimer 2 (Fig. 2b, c). The local resolution of the transmembrane helix core is around 3–3.5 Å in the cryo-EM map (Extended Data Fig. 5b). The secondary structural elements, as well as structural details of most amino acids, are well defined (Extended Data Fig. 6).

The cryo-EM structure of the ACAT1 tetramer is consistent with predictions from a previous biochemical study¹⁹. Each ACAT1 monomer has nine transmembrane helices, TM1–TM9 (Fig. 2c). The N-terminal portion of each of the ACAT1 monomers folds into a distinct four-helix bundle on the cytosolic side (Fig. 2a) and clearly mediates the tetramerization, in accordance with previous proposals¹⁹. Owing to the limited local resolution of the helices (Extended Data Fig. 5b), the identity of each helix could not be distinguished and they are not included in the final model. An ACAT1 variant with a N-terminal cytosolic helix deletion forms only a dimer¹⁹, suggesting that the cytosolic helix has a major role in the assembly of the tetramer. The dimer–dimer interface between the TMs of ACAT1-A and ACAT1-C comprises about 340 Å^2 . As well as the N-terminal cytosolic helices, residues Pro182, Val294 and Trp320 also contribute to tetramer assembly (Fig. 2d). Notably, two lipid molecules were observed at the interface of the two dimers. The density map of these two lipids was heterogeneous, indicating that it could comprise an average of several different lipids that bind to this

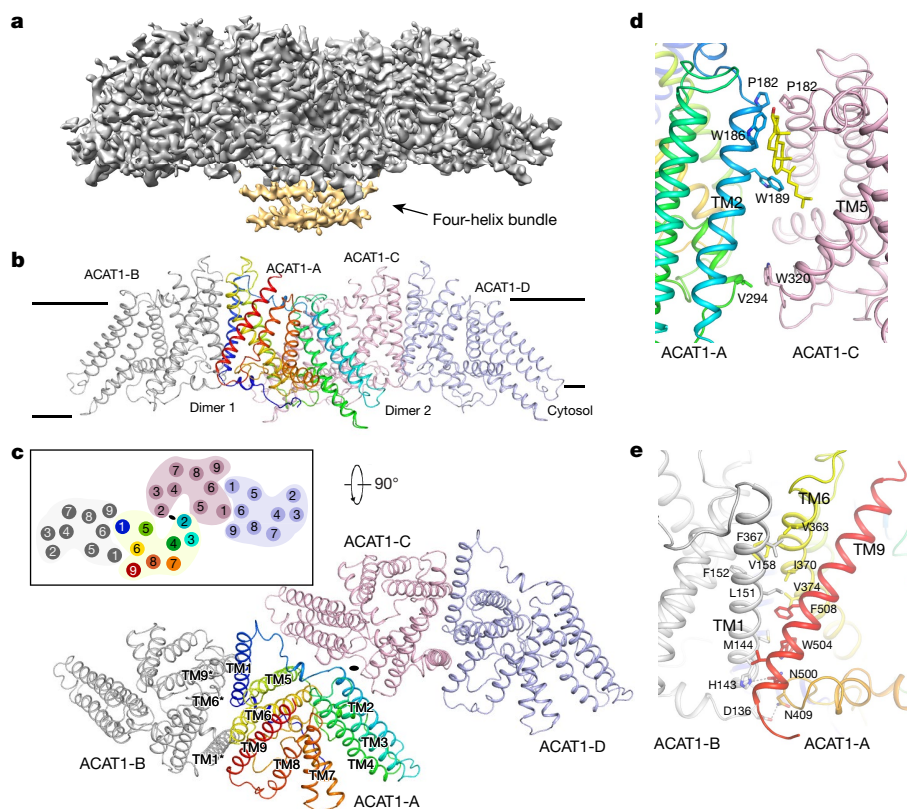


Fig. 2 | Overall structure of human ACAT1 holoenzyme.

a, Cryo-EM map of human ACAT1. The cytosolic four-helix bundle of ACAT1 is coloured in yellow. **b**, Overall structure of the ACAT1 tetramer, viewed from the side of the membrane. **c**, The top view of the holoenzyme showing the two dimers. TM1, TM6 and TM9 are located at the interface between the monomers and are involved in dimer assembly. **d**, Structural details of the interface between the dimers. The putative cholesterol molecule (yellow) is shown as a lipid representative. **e**, Structural details of the interface between the monomers. Residues that contribute to tetramer assembly, as well as those involved in interactions, are labelled.

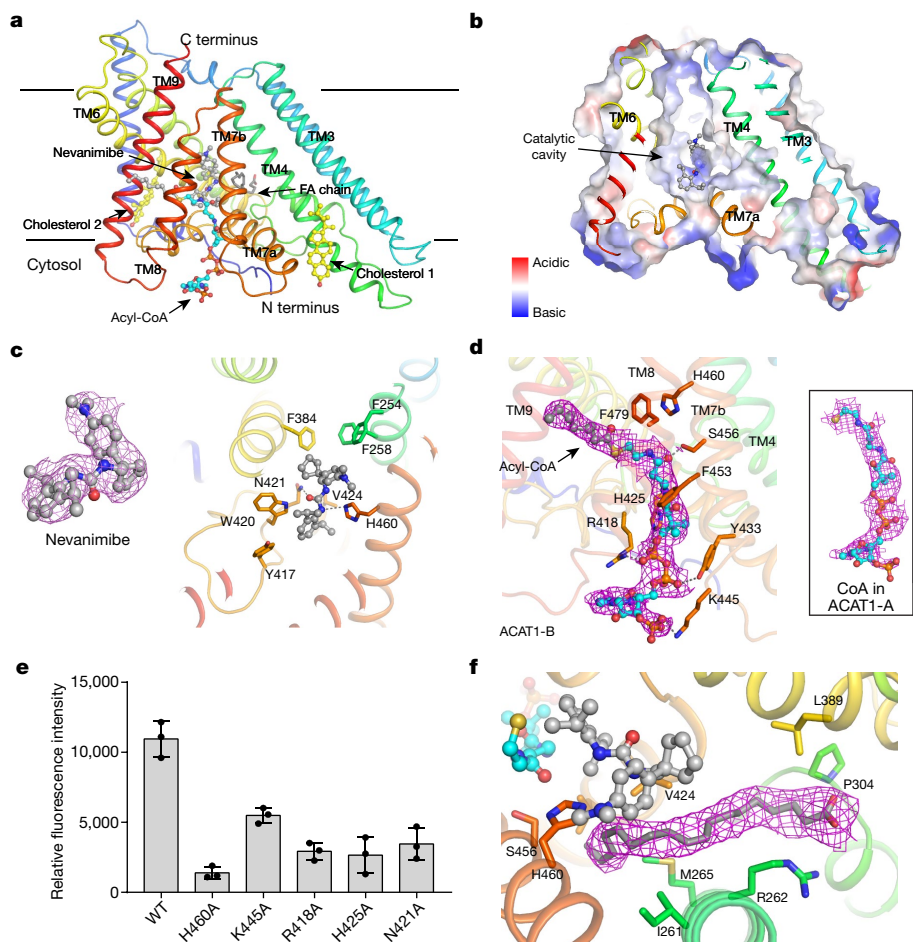


Fig. 3 | Overall structure of monomeric ACAT1 and its ligands. **a**, ACAT1 with nevanimibe (grey) and a fatty-acid (FA) chain (grey) in the hydrophobic cavity created by TM4–TM9, as well as one acyl-CoA (cyan) and two cholesterol molecules (yellow) on the cytosolic leaflet. **b**, Electrostatic surface representation of the catalytic cavity. **c**, The interaction of nevanimibe with residues of the catalytic cavity. **d**, Overall view of acyl-CoA with bound residues of ACAT1-B. **e**, Functional validation of the acyl-CoA binding site and the catalytic residues. Data are mean \pm s.d. ($n = 3$ technically independent experiments). Each experiment was reproduced at least twice on separate occasions with similar results. **f**, Details of the interaction between the fatty-acid chain and the residues of the catalytic cavity. Cryo-EM maps of nevanimibe, acyl-CoA, CoA and the fatty-acid chain are contoured at the 5σ , 5σ , 5σ and 7σ level, respectively. Residues are represented as sticks, and the dashed line represents hydrophilic interactions.

site. On the basis of the best fit using the cryo-EM map, we interpreted both of these densities as cholesterol (Fig. 2d); however, the presence of other lipids in these positions cannot be excluded.

The dimer interface between ACAT1-A and ACAT1-B comprises about $1,850 \text{ \AA}^2$ (Fig. 2e). TM1, TM6 and TM9 are involved in the assembly of the dimer. Although no symmetry was assumed within the dimer during cryo-EM data processing, ACAT1-A and ACAT1-B share nearly identical conformations with a root-mean-square deviation of 0.349. We then compared the structure of ACAT1 with the only other structurally characterized MBOAT—the prokaryotic enzyme DltB, which contains 11 TMs. Although no sequence identity has been detected between ACAT1 and DltB, TM4–TM11 of DltB show a topology similar to that of TM2–TM9 of ACAT1 (Extended Data Fig. 7a). Moreover, the putative catalytic His460 of ACAT1 aligns well with His336, the catalytic residue of DltB (Extended Data Fig. 7b).

Architecture of the catalytic cavity

ACAT1 contains a hydrophobic cavity formed by TM4–TM9, which is predicted to lie in the leaflet of the ER membrane. Several small-molecule densities are present in this cavity (Fig. 3a, b), including nevanimibe, which forms polar contacts with the catalytic His460⁸ (Fig. 3c). Nevanimibe is stabilized by Trp420 through a π – π interaction with an interplanar distance of 4 Å (Fig. 3c). The carbonyl oxygen of nevanimibe probably forms a hydrogen bond with Asn421, and the residues Phe254, Phe258, Phe384, Tyr417 and Val 424 are also involved in binding to nevanimibe (Fig. 3c). The position of nevanimibe suggests that it inhibits ACAT activity by sterically blocking substrate access to the catalytic His460 residue. This is reminiscent of the inhibition of sterol reductase in the biosynthetic pathway of

cholesterol²², in which the inhibitors directly block the entry of the substrate, sterol.

In ACAT1-B and ACAT1-D, the acyl-CoA substrate is found in the cytosolic site, and its fatty acid chain extends through TM8 and TM9 to reach the centre of the lipid bilayer (Fig. 3d). A similar density is observed in ACAT1-A and ACAT1-C, with a shorter length that perfectly fits CoA, one of the products of cholesterol esterification (Fig. 3d). Residues Arg418, His425, Tyr433, Lys445 and Ser456—which lie on the cytosolic leaflet of the protein—all have hydrophilic interactions with acyl-CoA, whereas Phe453 and Phe479 contribute hydrophobic contacts (Fig. 3d). 3'-Phosphoadenosine of CoA has few contacts with the cytosolic residues of ACAT1; this ensures the energetically favourable release of CoA from its binding site after transfer of the fatty acid to cholesterol.

The mutants ACAT1(F453C) and ACAT1(F479C)—in which the phenylalanine residues at positions 453 and 479, respectively, are substituted for cysteines—were previously reported to be inactive²³. Previous studies have also shown that Ser456 is required for activity²⁴. Our mutagenesis analysis confirmed that Arg418, His425 and Lys445 are required for ACAT1 activity in vitro (Fig. 3e, Extended Data Fig. 8). These findings support our structural observation that the substrate acyl-CoA enters the catalytic core through the cytosolic leaflet. In the current structure, the tail of acyl-CoA is pushed away from the catalytic core by nevanimibe. When nevanimibe is removed from the catalytic cavity, the lack of steric hindrance enables the thioester bond of acyl-CoA to interact with His460. Our mutagenesis data show that Asn421, which binds to nevanimibe, is also required for the esterification of cholesterol (Fig. 3e). This implies that Asn421 may be involved in stabilizing the reaction intermediate.

As well as acyl-CoA, a fatty-acid-like density is present in the catalytic core (Fig. 3a, f). We interpreted this density as oleic acid; however, the

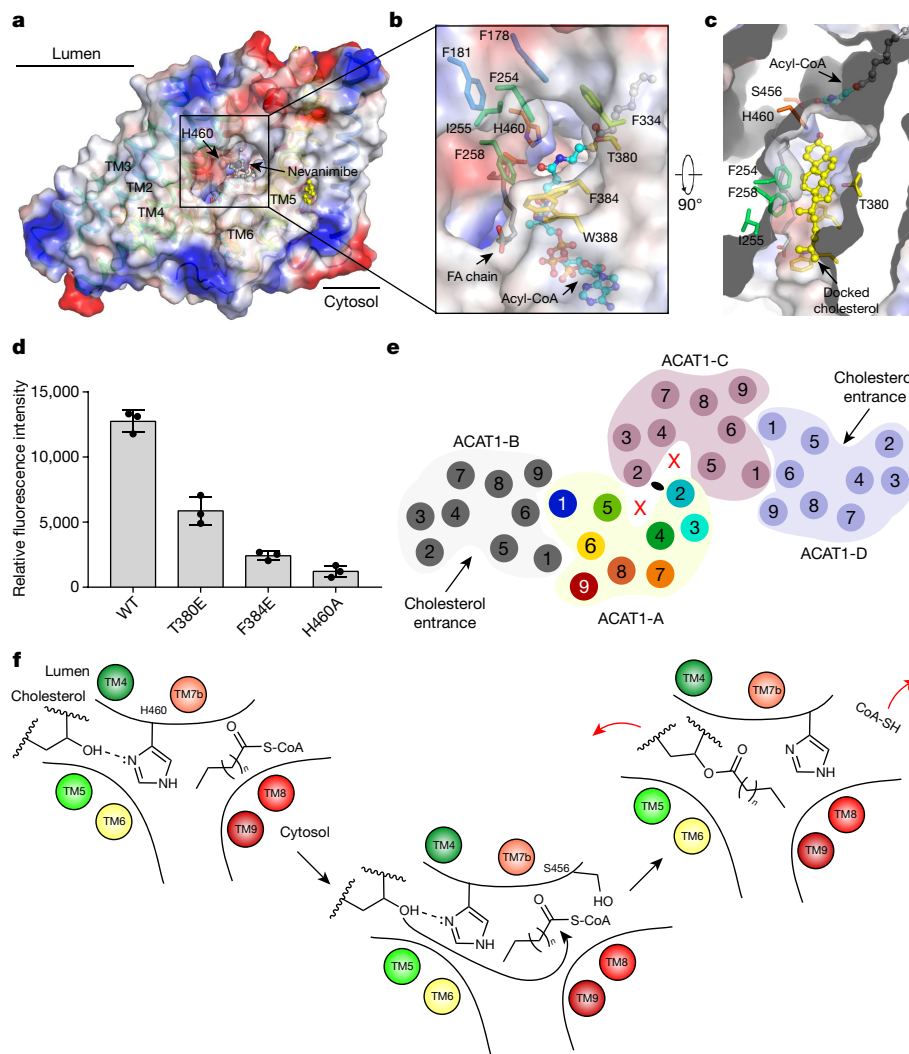


Fig. 4 | The cholesterol entrance. **a**, Electrostatic surface representation of the cavity in the luminal leaflet of the membrane. His460 is blocked by nevanimibe in the cavity. **b**, A magnified view of the putative entrance for cholesterol from the membrane cavity. His460 is exposed after the removal of nevanimibe. **c**, Top view of the catalytic cavity after docking of cholesterol using PatchDock²⁵. **d**, Functional validation of the cholesterol entrance. Data

are mean \pm s.d. ($n = 3$ technically independent experiments). Each experiment was reproduced at least twice on separate occasions with similar results. **e**, The tetrameric conformation of ACAT1 prevents the access of cholesterol substrates into the cavities of ACAT1-A and ACAT1-C. **f**, Proposed reaction mechanism for the esterification of cholesterol, showing the putative reaction substrates and the product, cholesteryl ester.

presence of other fatty acids in this position cannot be excluded. One end of this fatty acid chain is within 4–5 Å of both the nevanimibe molecule and the catalytic His460. The residues Ile261, Met265, Pro304, Leu389, Val424 and Leu428 make numerous hydrophobic contacts with this chain. Arg262 is located close to the putative carboxylate group of the fatty acid chain, and seemingly detains the fatty acid chain in the cavity (Fig. 3f). We speculate that this fatty acid molecule can render the membrane cavity sufficiently rigid to enable cholesterol to access the catalytic His460.

Working model of ACAT1

The structure also reveals a hydrophobic tunnel—created by residues in TM2, TM4, TM5 and TM6—that is open to the luminal leaflet of the ER membrane (Fig. 4a). Nevanimibe blocks the catalytic residue His460, which is located in the centre of the tunnel, preventing access to it from the ER membrane (Fig. 4a). When nevanimibe is subtracted from the structure, His460 becomes exposed to the membrane environment (Fig. 4b). Docking results from PatchDock²⁵ show that cholesterol could fit into this tunnel; in particular, the 3 β -hydroxyl group of cholesterol

can reach His460, where the esterification could be initiated (Fig. 4c). To verify this docking result, we generated two mutants—ACAT1(T380E) and ACAT1(F384E)—to prevent the entrance of cholesterol. Activity assays with these mutants supported our hypothesis (Fig. 4d, Extended Data Fig. 8). Notably, in the tetrameric state, the cholesterol entrance sites in ACAT1-A and ACAT1-C are blocked by the dimer–dimer interface (Fig. 4e), and cholesterol can access the catalytic core only through the tunnel formed by ACAT1-B and ACAT1-D. The blocking of cholesterol entry by the dimer–dimer interface might explain previous findings that the V_{\max} of the ACAT tetramer is much lower than that of the dimer¹⁹.

On the basis of these observations, we propose a working model of ACAT1-mediated cholesterol esterification (Fig. 4f). Acyl-CoA enters the catalytic cavity through the cytosolic site (Fig. 3a) and cholesterol enters through the luminal leaflet tunnel (Fig. 4a). His460 acts as a base to deprotonate the 3 β -hydroxyl group of cholesterol, and acyl-CoA can be stabilized through interactions with the residues located in the cytosolic site (Fig. 3d). This reactive cholesterol intermediate then attacks acyl-CoA to form the cholesteryl ester product. Residues in the catalytic core—such as Ser456 and Asn421—may stabilize the intermediate product, as biochemical analysis shows that mutations of

these residues can abolish the reaction²⁴ (Fig. 3e); however, technical limitations have so far prevented us from specifying the role of each of these residues. The ester product is released into the ER membrane via the same pore through which cholesterol entered. As cholesteryl esters accumulate in the lipid bilayer, the bilayer splits, forming lipid droplets that are coated with a phospholipid monolayer. The close juxtaposition of two substrates—cholesterol and the acyl chain—in ACAT1 is reminiscent of the active site of sterol reductase in the cholesterol biosynthetic pathway, in which the substrate, sterol, is adjacent to the reducing cofactor NADPH in the central cavity²⁶.

On the basis of biochemical assays, it has been suggested that cholesterol is the most efficient allosteric activator of ACAT1²⁷. We observed a cholesterol molecule in our structure, located in a hydrophobic pocket surrounded by TM3, TM4 and TM7a in the cytosolic leaflet (Extended Data Fig. 9a). We suggest that this additional cholesterol molecule may stabilize TM4 and TM7a in a configuration that enables the substrate cholesterol molecule to enter the tunnel (Fig. 4c). A second cholesterol molecule in our structure is located in a hydrophobic pocket flanked by TM1, TM5 and TM6 (Extended Data Fig. 9b). The sterol ring makes a hydrophobic contact with Trp408 that might affect the conformation of Trp420 and Asn421 in the catalytic core. This interaction would maintain TM6 in a conformation that enables the substrate cholesterol molecule to enter the hydrophobic tunnel (Fig. 4c). Notably, this second cholesterol is located at the interface of ACAT1-A and ACAT1-B, and may also have a role in stabilizing the ACAT1 dimer (Extended data Fig. 9c). To validate the physiological importance of these two putative allosteric sites, we introduced two mutations on each site and measured the activity of the resultant enzymes. The mutation of Phe382 and Trp408 led to a 90% loss of activity compared with the wild type, whereas the mutation of Arg272 and Trp438 did not affect the activity (Extended Data Figs. 8, 9d). These data suggest that the cholesterol-binding site among TM1, TM5 and TM6 (Extended Data Fig. 9c) has a primary role in the cholesterol-mediated activation of ACAT1. Further investigation will be required in order to fully elucidate the mechanism of allosteric activation.

Discussion

We report the cryo-EM structure of human ACAT1 with its inhibitor nevanimibe. ACAT1 was captured in an intermediate state in which several small molecules are present surrounding the active site, including the substrate acyl-CoA and the inhibitor nevanimibe. These findings help us to understand the catalysis of cholesterol esterification and also reveal the mode of action of the ACAT inhibitor nevanimibe.

Mammalian cells obtain cholesterol through the LDL-receptor-mediated uptake of LDL particles. Niemann–Pick C1 Protein (NPC1) and NPC2 collaborate to export cholesterol from the lysosomal membrane and shuttle it to the ER^{28–30}. The concentration of cholesterol in the ER is crucial for activation of the SREBP pathway⁹: if the concentration of cholesterol exceeds 5%, the SREBP pathway will be turned off to prevent cholesterol biosynthesis and LDL-derived cholesterol uptake. In the monitoring of cholesterol concentration by this pathway, free cholesterol first interacts with the SREBP cleavage-activating protein SCAP, an ER cholesterol sensor⁹. ACAT1 then senses free cholesterol through its allosteric site. At low cholesterol concentrations, ACAT1 will not efficiently catalyse esterification; however, at high concentrations, excess cholesterol can allosterically promote esterification. This mechanism would ensure that ACAT1 activity can be regulated by the concentration of free cholesterol in the membrane, to maintain cholesterol homeostasis in the ER.

Online content

Any methods, additional references, Nature Research reporting summaries, source data, extended data, supplementary information,

acknowledgements, peer review information; details of author contributions and competing interests; and statements of data and code availability are available at <https://doi.org/10.1038/s41586-020-2295-8>.

- Radhakrishnan, A., Goldstein, J. L., McDonald, J. G. & Brown, M. S. Switch-like control of SREBP-2 transport triggered by small changes in ER cholesterol: a delicate balance. *Cell Metab.* **8**, 512–521 (2008).
- Rogers, M. A. et al. Acyl-CoA:cholesterol acyltransferases (ACATs/SOATs): enzymes with multiple sterols as substrates and as activators. *J. Steroid Biochem. Mol. Biol.* **151**, 102–107 (2015).
- Goldstein, J. L. & Brown, M. S. The LDL receptor. *Arterioscler. Thromb. Vasc. Biol.* **29**, 431–438 (2009).
- Hutter-Paier, B. et al. The ACAT inhibitor CP-113,818 markedly reduces amyloid pathology in a mouse model of Alzheimer's disease. *Neuron* **44**, 227–238 (2004).
- Yagyu, H. et al. Absence of ACAT-1 attenuates atherosclerosis but causes dry eye and cutaneous xanthomatosis in mice with congenital hyperlipidemia. *J. Biol. Chem.* **275**, 21324–21330 (2000).
- Yang, W. et al. Potentiating the antitumour response of CD8⁺ T cells by modulating cholesterol metabolism. *Nature* **531**, 651–655 (2016).
- LaPensee, C. R. et al. ATR-101, a selective and potent inhibitor of acyl-CoA acyltransferase 1, induces apoptosis in H295R adrenocortical cells and in the adrenal cortex of dogs. *Endocrinology* **157**, 1775–1788 (2016).
- Guo, Z. Y., Lin, S., Heinen, J. A., Chang, C. C. & Chang, T. Y. The active site His-460 of human acyl-coenzyme A:cholesterol acyltransferase 1 resides in a hitherto undisclosed transmembrane domain. *J. Biol. Chem.* **280**, 37814–37826 (2005).
- Brown, M. S., Radhakrishnan, A. & Goldstein, J. L. Retrospective on cholesterol homeostasis: the central role of Scap. *Annu. Rev. Biochem.* **87**, 783–807 (2018).
- Porter, F. D. & Herman, G. E. Malformation syndromes caused by disorders of cholesterol synthesis. *J. Lipid Res.* **52**, 6–34 (2011).
- Xu, Y., Du, X., Turner, N., Brown, A. J. & Yang, H. Enhanced acyl-CoA:cholesterol acyltransferase activity increases cholesterol levels on the lipid droplet surface and impairs adipocyte function. *J. Biol. Chem.* **294**, 19306–19321 (2019).
- Masumoto, N. et al. Membrane bound O-acyltransferases and their inhibitors. *Biochem. Soc. Trans.* **43**, 246–252 (2015).
- Hofmann, K. A superfamily of membrane-bound O-acyltransferases with implications for Wnt signaling. *Trends Biochem. Sci.* **25**, 111–112 (2000).
- Ma, D. et al. Crystal structure of a membrane-bound O-acyltransferase. *Nature* **562**, 286–290 (2018).
- Chang, C. C., Huh, H. Y., Cadigan, K. M. & Chang, T. Y. Molecular cloning and functional expression of human acyl-coenzyme A:cholesterol acyltransferase cDNA in mutant Chinese hamster ovary cells. *J. Biol. Chem.* **268**, 20747–20755 (1993).
- Anderson, R. A. et al. Identification of a form of acyl-CoA:cholesterol acyltransferase specific to liver and intestine in nonhuman primates. *J. Biol. Chem.* **273**, 26747–26754 (1998).
- Cases, S. et al. ACAT-2, a second mammalian acyl-CoA:cholesterol acyltransferase. Its cloning, expression, and characterization. *J. Biol. Chem.* **273**, 26755–26764 (1998).
- Oelkers, P., Behari, A., Cromley, D., Billheimer, J. T. & Sturley, S. L. Characterization of two human genes encoding acyl coenzyme A:cholesterol acyltransferase-related enzymes. *J. Biol. Chem.* **273**, 26765–26771 (1998).
- Yu, C. et al. Role of the N-terminal hydrophilic domain of acyl-coenzyme A:cholesterol acyltransferase 1 on the enzyme's quaternary structure and catalytic efficiency. *Biochemistry* **41**, 3762–3769 (2002).
- Nissen, S. E. et al. Effect of ACAT inhibition on the progression of coronary atherosclerosis. *N. Engl. J. Med.* **354**, 1253–1263 (2006).
- Cao, J. et al. Targeting Acyl-CoA:diacylglycerol acyltransferase 1 (DGAT1) with small molecule inhibitors for the treatment of metabolic diseases. *J. Biol. Chem.* **286**, 41838–41851 (2011).
- Long, T. et al. Structural basis for human sterol isomerase in cholesterol biosynthesis and multidrug recognition. *Nat. Commun.* **10**, 2452 (2019).
- Guo, Z. Y., Chang, C. C. & Chang, T. Y. Functionality of the seventh and eighth transmembrane domains of acyl-coenzyme A:cholesterol acyltransferase 1. *Biochemistry* **46**, 10063–10071 (2007).
- Das, A., Davis, M. A. & Rudel, L. L. Identification of putative active site residues of ACAT enzymes. *J. Lipid Res.* **49**, 1770–1781 (2008).
- Schneidman-Duhovny, D., Inbar, Y., Nussinov, R. & Wolfson, H. J. PatchDock and SymmDock: servers for rigid and symmetric docking. *Nucleic Acids Res.* **33**, W363–W367 (2005).
- Li, X., Roberti, R. & Blobel, G. Structure of an integral membrane sterol reductase from *Methylobacterium alcaliphilum*. *Nature* **517**, 104–107 (2015).
- Liu, J., Chang, C. C., Westover, E. J., Covey, D. F. & Chang, T. Y. Investigating the allostereism of acyl-CoA:cholesterol acyltransferase (ACAT) by using various sterols: in vitro and intact cell studies. *Biochem. J.* **391**, 389–397 (2005).
- Li, X., Saha, P., Li, J., Blobel, G. & Pfeffer, S. R. Clues to the mechanism of cholesterol transfer from the structure of NPC1 middle luminal domain bound to NPC2. *Proc. Natl Acad. Sci. USA* **113**, 10079–10084 (2016).
- Pfeffer, S. R. NPC intracellular cholesterol transporter 1 (NPC1)-mediated cholesterol export from lysosomes. *J. Biol. Chem.* **294**, 1706–1709 (2019).
- Long, T. et al. Structural basis for itraconazole-mediated NPC1 inhibition. *Nat. Commun.* **11**, 152 (2020).

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

© The Author(s), under exclusive licence to Springer Nature Limited 2020

Article

Methods

Data reporting

No statistical methods were used to predetermine sample size. The experiments were not randomized and the investigators were not blinded to allocation during experiments and outcome assessment.

Protein expression and purification

The cDNA of human ACAT1 (GenBank: BC028940.1) was cloned into pEG BacMam with a C-terminal Flag-tag. The protein was expressed using baculovirus-mediated transduction of mammalian HEK-293S GnT1⁻ cells (ATCC) that were routinely monitored for mycoplasma contamination. For nevanimibe-bound proteins, 10 μ M ligand was added in all purification steps. At 48 h after infection, the cells were disrupted by sonication in buffer A, containing 20 mM HEPES, pH 7.5, 150 mM NaCl with 1 mM phenylmethylsulfonyl fluoride and 5 μ g ml⁻¹ leupeptin. After low-speed centrifugation, the resulting supernatant was incubated in buffer B with 1% (w/v) lauryl maltose neopentyl glycol (LMNG, Anatrace) for 1 h at 4 °C. The lysate was centrifuged again, and the supernatant was loaded onto a Flag-M2 affinity column (Sigma-Aldrich). After washing twice, the protein was eluted in 20 mM HEPES, pH 7.5, 150 mM NaCl, 100 μ g ml⁻¹ 3 × Flag peptide, 0.01% LMNG or 0.06% digitonin (Acros Organics) and concentrated. The concentrated protein was purified by size-exclusion chromatography (SEC) on a Superose 6 Increase column (GE Healthcare) in a buffer containing buffer A and 0.06% (w/v) digitonin for cryo-EM study (Extended Data Fig. 2) or 1% CHAPS (Anatrace) for the fluorescence activity assay (Extended Data Fig. 8). Point mutations were introduced into the coding region of ACAT1 by site-directed mutagenesis using the QuikChange II XL Site-Directed Mutagenesis Kit (Agilent Technologies). The coding region of each plasmid was sequenced to ensure the integrity of the construct.

Fluorescence-based ACAT1 assay

ACAT activity was measured by monitoring the CoA released from an acyltransferase-mediated reaction²¹. The sulfhydryl (-SH) group of CoA can react with CPM and the resulting highly fluorescent product is readily detected. Mixed micelles with 2 mM cholesterol/10 mM POPC/18.6 mM taurocholate (Alfa Aesar) in reaction buffer containing 100 mM HEPES pH 7.5 and 150 mM NaCl were prepared as described previously³¹. The assay was carried out in a total volume of 20 μ l under the following conditions: 15 μ l mixed micelles, 2.5 μ l protein (at 7.2 μ M concentration) and 2.5 μ l 400 μ M oleoyl-CoA (MP Biomedicals). The reaction was initiated by the addition of oleoyl-CoA and incubated at 37 °C for 3 min (Fig. 1c) or the indicated time (Fig. 1b). For the functional validations (Figs. 3e, 4d, Extended Data Fig. 9d), the assay was carried out in a total volume of 10 μ l under the following conditions: 7 μ l mixed micelles, 2.6 μ l protein (at a concentration of 1.14 μ M) and 0.4 μ l 1.25 mM oleoyl-CoA. The reaction was initiated by the addition of protein and incubated at 37 °C for 60 min. The reaction was terminated by adding 3 μ l 10% SDS. Then 90 μ l of 50 μ M CPM in reaction buffer was added to the reaction system and the mixture was transferred to a 96-well plate. The plate was incubated at room temperature for 30 min, followed by detection of the fluorescent signal using a BioTek Synergy Neo2 Hybrid Multi-Mode Reader (excitation 355 nm; emission 460 nm). Relative fluorescence intensity was obtained by subtracting the fluorescence intensity of the oleoyl-CoA-free reaction system for the corresponding protein. To measure the IC₅₀, different concentrations of nevanimibe as indicated were added to the mixed micelles, then the protein was added and the mixture was incubated at 37 °C for 3 min. The reaction was started and terminated as described above. IC₅₀ calculations were performed using GraphPad Prism7.

EM sample preparation, imaging and processing for 200 kV cryo-TEM

For 200 kV cryo-transmission electron microscopy (cryo-TEM), the nevanimibe-bound ACAT1 sample was crosslinked by the addition of

0.1% glutaraldehyde (Sigma-Aldrich) and incubated at room temperature for 30 min. The reaction was then terminated by the addition of 50 mM Tris, pH 8.0 at room temperature for 10 min. The cross-linked sample was further purified by SEC on a Superose 6 Increase column in a buffer containing buffer A and 0.06% (w/v) digitonin (Extended Data Fig. 3a). The sample (4 mg ml⁻¹) was applied to Quantifoil R1.2/1.3 400 mesh Au holey carbon grids (Quantifoil). The grids were then blotted and plunged into liquid ethane for flash freezing using a Vitrobot Mark IV (FEI). The grids were imaged in a 200 kV Talos Arctica (FEI) with a Gatan K3 Summit direct electron detector (Gatan). Data were collected at 0.89 Å per pixel with a dose rate of 32 electrons per physical pixel per second. Images were recorded for 1.5 s exposures in 50 subframes with a total dose of 60 e⁻ Å⁻².

Motion correction was performed using the program MotionCor2³². The contrast transfer function (CTF) was estimated using CTFFIND4³³. To generate ACAT1 templates for automatic picking, around 3,000 particles were manually picked and classified by 2D classification in RELION-3³⁴. After auto-picking, the low-quality images and false-positive particles were removed manually. The remaining 165,508 particles were extracted for subsequent 2D classification. A low-resolution cryo-EM map of ACAT1, which was generated from 3,200 particles by RELION-3, was used as the initial model for 3D classification. The best class—containing 70,964 particles—was selected for the initial 3D-refinement, followed by Bayesian polishing and CTF refinement with beam tilt correction in RELION-3. The resulting particles were used for the final 3D-refinement with a soft mask, and solvent-flattened Fourier shell correlations (FSCs) yielded a reconstruction at 8.3 Å revealing clear secondary structural elements. The resolution was estimated using post-processing with the FSC criteria of 0.143.

EM sample preparation and imaging for 300 kV cryo-TEM

The nevanimibe-bound ACAT1 sample (4 mg ml⁻¹ native protein, not cross-linked) was applied to Quantifoil R1.2/1.3 300 or 400 mesh Au holey carbon grids (Quantifoil). The grids were then blotted and plunged into liquid ethane for flash freezing using a Vitrobot Mark IV (FEI). The grids were imaged in a 300 kV Titan Krios (FEI) with a Gatan K3 Summit direct electron detector (Gatan). Data were collected at 0.833 Å per pixel with a dose rate of 23 electrons per physical pixel per second. Images were recorded for 1.8 s exposures in 60 subframes with a total dose of 60 e⁻ Å⁻².

Imaging processing and 3D reconstruction for 300 kV cryo-TEM

The images were collected in two sessions (4,124 images from 400-mesh, 5,180 images from 300-mesh Au holey carbon grids; Extended Data Fig. 4). Dark-subtracted images collected at super-resolution mode were first normalized by gain reference and binned twofold, which resulted in a pixel size of 0.833 Å. Motion correction was performed using the program MotionCor2³². The CTF was estimated using CTFFIND4³³. To generate ACAT1 templates for automatic picking, around 3,000 particles were manually picked and classified by 2D classification in RELION-3³⁴. After auto-picking, low-quality images and false-positive particles were removed manually. About 1.5 million/1.9 million (from 400-mesh/300-mesh grid) particles of ACAT1 with nevanimibe were extracted. We used the cryo-EM structure of human ACAT1 that we determined, with the data collected from a 200 kV Arctica (FEI) low-pass filtered to 40 Å as the initial model with a C2 symmetry for 3D classification. The best class, containing 305,921/373,608 particles, provided a 6.1 Å/6.8 Å map after 3D auto-refinement without a mask in RELION-3.

For the particles from the 400-mesh grid, CTF refinement and Bayesian polishing of particles were then performed using RELION-3. The 3D refinement using a soft mask and solvent-flattened FSCs yielded a reconstruction at 4.3 Å. Then, a 3D classification without image alignment was performed. The resulting 242,501 particles were refined using a soft mask and solvent-flattened FSCs yielded a reconstruction at 4.23 Å. For the particles from the 300-mesh grid, Bayesian polishing was then

performed using RELION-3. The 3D refinement using a soft mask and solvent-flattened FSCs yielded a reconstruction at 5.8 Å. Then, a 2D classification was performed. The resulting 240,113 particles refined without a mask yielded a reconstruction at 5.7 Å. Finally, we combined these two datasets with a total of 482,614 particles and performed a masked 3D classification without image alignment. The best class, including 263,839 particles refined using a soft mask and solvent-flattened FSCs, yielded a reconstruction at 4.1 Å. Applying a soft mask in RELION-3 post-processing yielded a final cryo-EM map of 4.1 Å. Two rounds of CTF refinements were performed, and the resulting particles refined using a soft mask and solvent-flattened FSCs yielded a reconstruction at 3.71 Å. Applying a soft mask in RELION-3 post-processing yielded a final cryo-EM map of 3.67 Å. Resolution was estimated using the FSC 0.143 criterion.

Model construction

To obtain better side-chain densities for model building, we sharpened the map using post-processing in RELION-3 with a B-factor value of -150 Å^2 . The initial model was built using phenix.map_to_model³⁵ and then manually adjusted using Coot³⁶. Large aromatic or hydrophobic residues were assigned to facilitate the register of the transmembrane helices. The densities of residues 1–117 of human ACAT1 were neither resolved nor built. Residues 285–288 and 351–357 were built with alanine owing to limited local resolution.

Model refinement and validation

The model was refined in real space using PHENIX⁴¹ and also in reciprocal space using Refmac with secondary-structure restraints and stereochemical restraints³⁷. For cross-validations, the final model was refined against one of the half maps generated by 3D auto-refine and the model versus map FSC curves were generated in the Comprehensive validation module in PHENIX. PHENIX and MolProbity³⁸ were used to validate the final model. Local resolutions were estimated using ResMap³⁹. Structure figures were generated using PyMOL (<http://www.pymol.org>) and Chimera⁴⁰.

Reporting summary

Further information on research design is available in the Nature Research Reporting Summary linked to this paper.

Data availability

The 3D cryo-EM density map has been deposited in the Electron Microscopy Data Bank under the accession number EMD-21390. Atomic

coordinates for the atomic model have been deposited in the Protein Data Bank under the accession number 6VUM. Source Data for Figs. 1, 3, 4 and Extended Data Fig. 9 are provided with the paper. All other data are available from the corresponding authors upon reasonable request.

31. Chang, C. C. et al. Recombinant acyl-CoA:cholesterol acyltransferase-1 (ACAT-1) purified to essential homogeneity utilizes cholesterol in mixed micelles or in vesicles in a highly cooperative manner. *J. Biol. Chem.* **273**, 35132–35141 (1998).
32. Zheng, S. Q. et al. MotionCor2: anisotropic correction of beam-induced motion for improved cryo-electron microscopy. *Nat. Methods* **14**, 331–332 (2017).
33. Rohou, A. & Grigorieff, N. CTFFIND4: Fast and accurate defocus estimation from electron micrographs. *J. Struct. Biol.* **192**, 216–221 (2015).
34. Zivanov, J. et al. New tools for automated high-resolution cryo-EM structure determination in RELION-3. *eLife* **7**, e42166 (2018).
35. Terwilliger, T. et al. A fully automatic method yielding initial models from high-resolution cryo-electron microscopy maps. *Nat. Methods* **15**, 905–908 (2018).
36. Emsley, P. & Cowtan, K. Coot: model-building tools for molecular graphics. *Acta Crystallogr. D* **60**, 2126–2132 (2004).
37. Murshudov, G. N., Vagin, A. A. & Dodson, E. J. Refinement of macromolecular structures by the maximum-likelihood method. *Acta Crystallogr. D* **53**, 240–255 (1997).
38. Chen, V. B. et al. MolProbity: all-atom structure validation for macromolecular crystallography. *Acta Crystallogr. D* **66**, 12–21 (2010).
39. Kucukelbir, A., Sigworth, F. J. & Tagare, H. D. Quantifying the local resolution of cryo-EM density maps. *Nat. Methods* **11**, 63–65 (2014).
40. Pettersen, E. F. et al. UCSF Chimera—a visualization system for exploratory research and analysis. *J. Comput. Chem.* **25**, 1605–1612 (2004).
41. Adams, P. D. et al. PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr. D Biol. Crystallogr.* **66**, 213–221 (2010).

Acknowledgements We thank M. Brown and J. Goldstein for support throughout the project. Data were collected at the UT Southwestern Medical Center Cryo-EM Facility (funded in part by the CPRIT Core Facility Support Award RP170644). We thank D. Stoddard for assistance with data collection; R. DeBose-Boyd, L. Esparza and L. Friedberg for technical support; E. Coutavas and E. Debler for discussion; and A. Lemoff at the UT Southwestern Proteomics Core for mass spectrometry identification. This work was supported by National Institutes of Health grants P01 HL020948, R01 GM134700 and R01 GM135343; the Endowed Scholars Program in Medical Science of UT Southwestern Medical Center; O'Donnell Junior Faculty Funds; and Welch Foundation grant I-1957 to X.L. X.Q. is the recipient of a DDBrown Fellow of the Life Sciences Research Foundation. X.L. is a Damon Runyon-Rachleff Innovator supported by the Damon Runyon Cancer Research Foundation (DRR-53-19) and a Rita C. and William P. Clements Jr. Scholar in Biomedical Research at UT Southwestern Medical Center.

Author contributions X.L. conceived the project and designed the research with T.L., T.L. and A.H. performed the biochemical study. T.L., Y.S. and X.Q. carried out cryo-EM work, built the model and refined the structure. T.L., Y.S., X.Q. and X.L. analysed the data, and all authors contributed to the preparation of the manuscript. X.L. wrote the manuscript.

Competing interests The authors declare no competing interests.

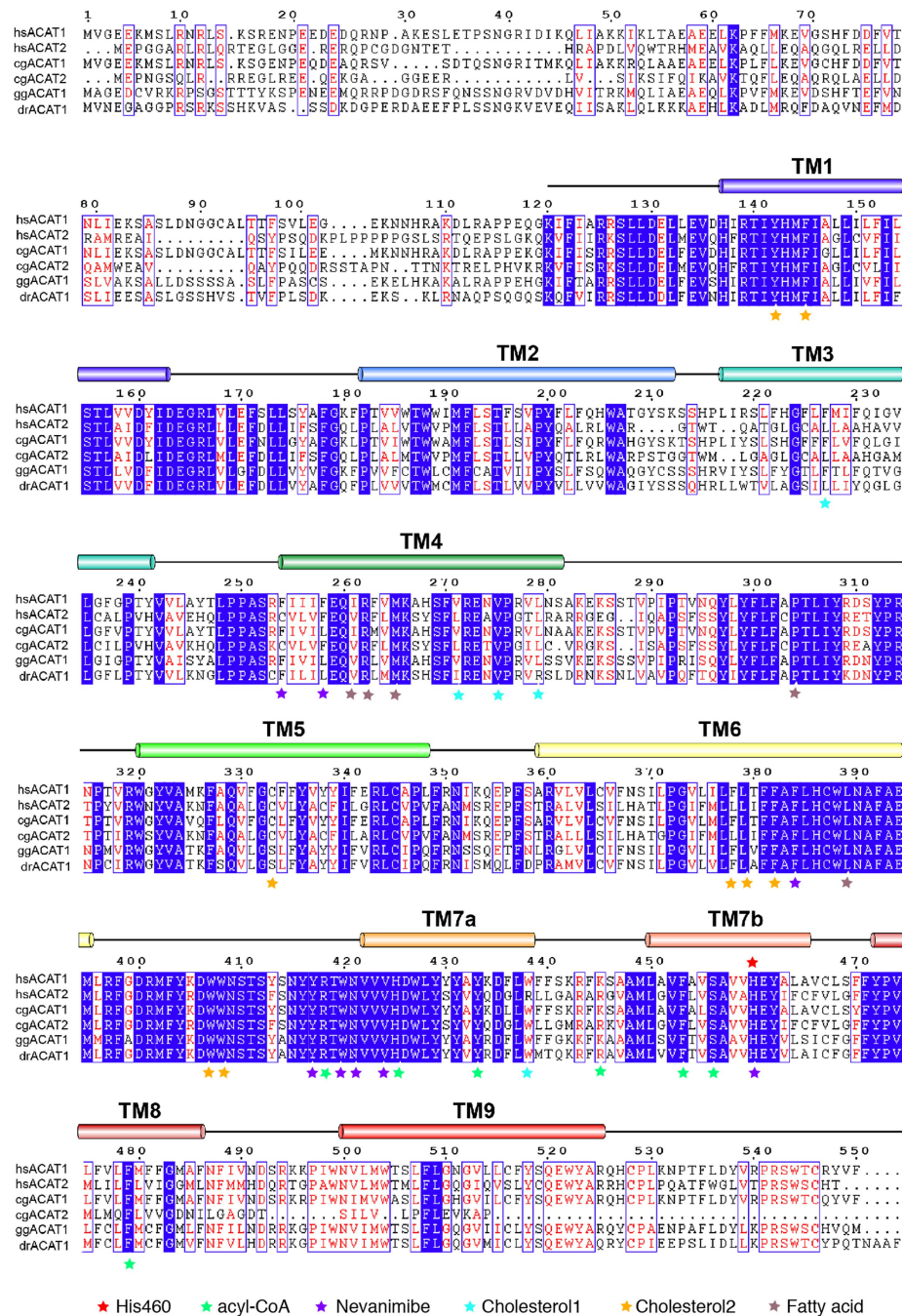
Additional information

Supplementary information is available for this paper at <https://doi.org/10.1038/s41586-020-2295-8>.

Correspondence and requests for materials should be addressed to X.L.

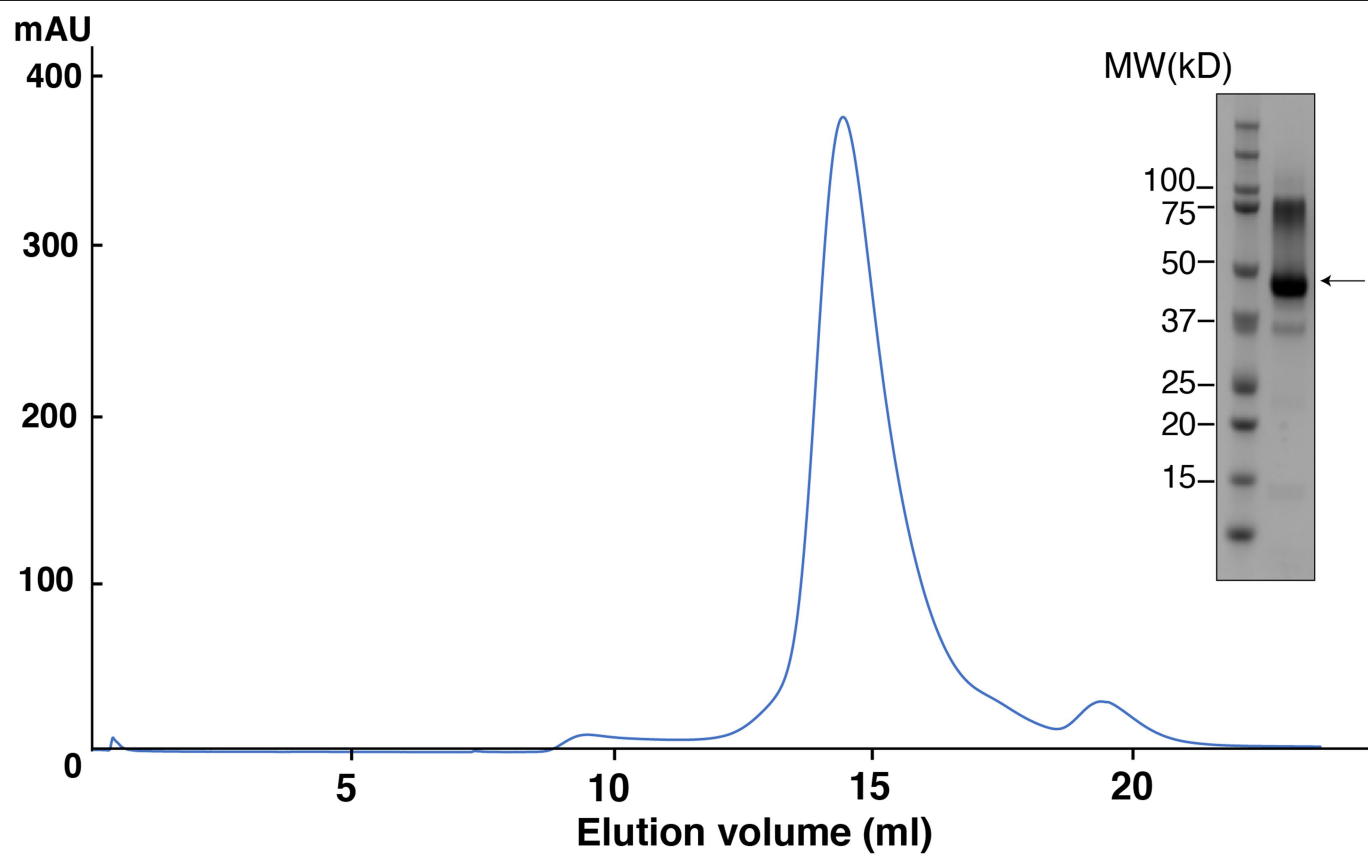
Peer review information Nature thanks David Drew, Savvas N. Savvides and Alan A. Tall for their contribution to the peer review of this work.

Reprints and permissions information is available at <http://www.nature.com/reprints>.



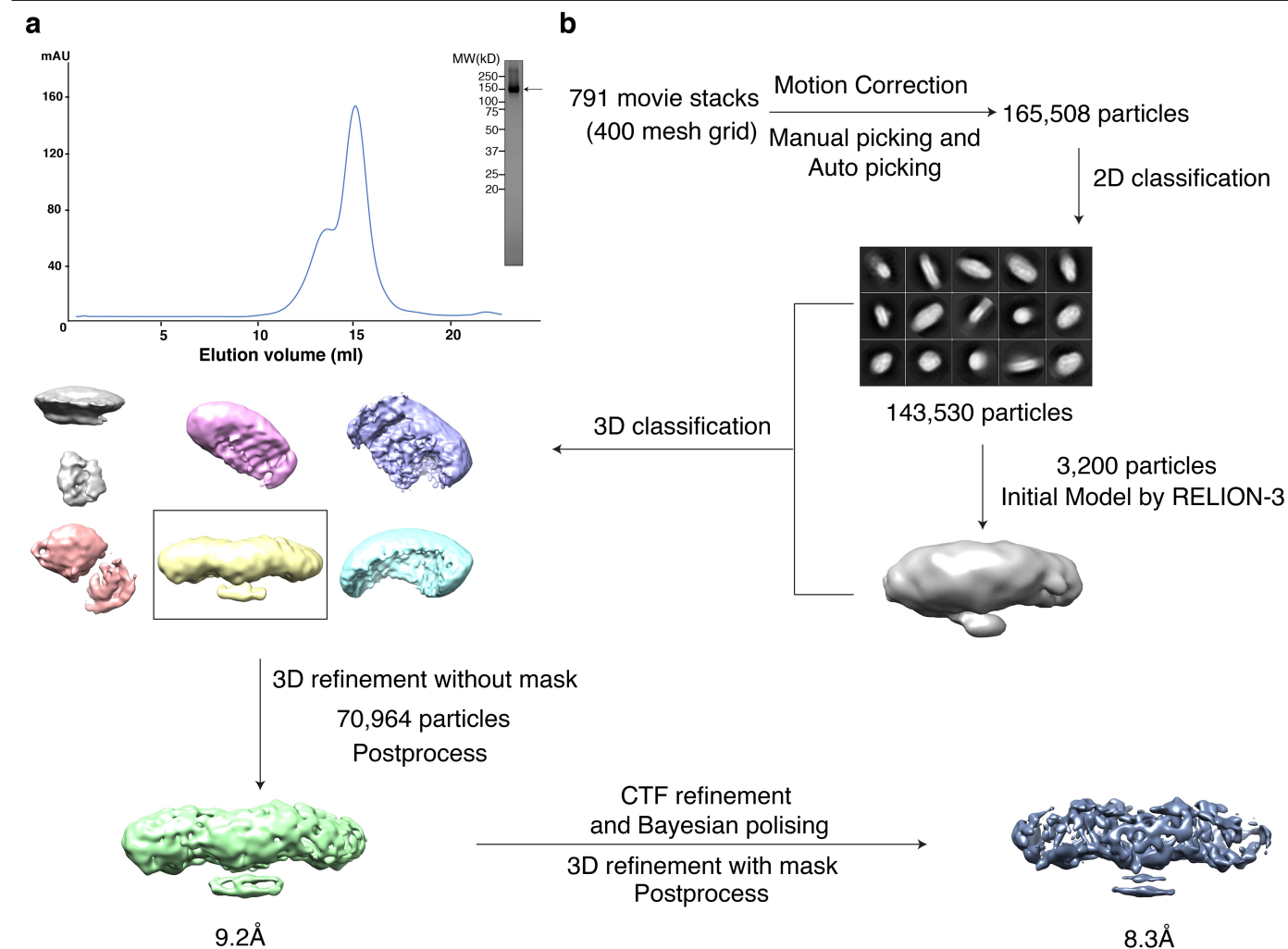
Extended Data Fig. 1 | Sequence alignment of human ACAT1 and ACAT2, hamster ACAT1 and ACAT2, chicken ACAT1 and zebrafish ACAT1. The transmembrane helices and the residue numbers of human ACAT1 are indicated above the protein sequence. The specific residues necessary for

ligand binding are labelled with an asterisk. Hs, *Homo sapiens* (human); cg, *Cricetulus griseus* (Chinese hamster); gg, *Gallus gallus* (chicken); dr, *Danio rerio* (zebrafish).



Extended Data Fig. 2 | Expression and purification of human ACAT1 with nevanimibe. Representative gel-filtration chromatogram of ACAT1 (Superose 6 Increase 10/30 column). The SDS-PAGE gel (with molecular weight markers)

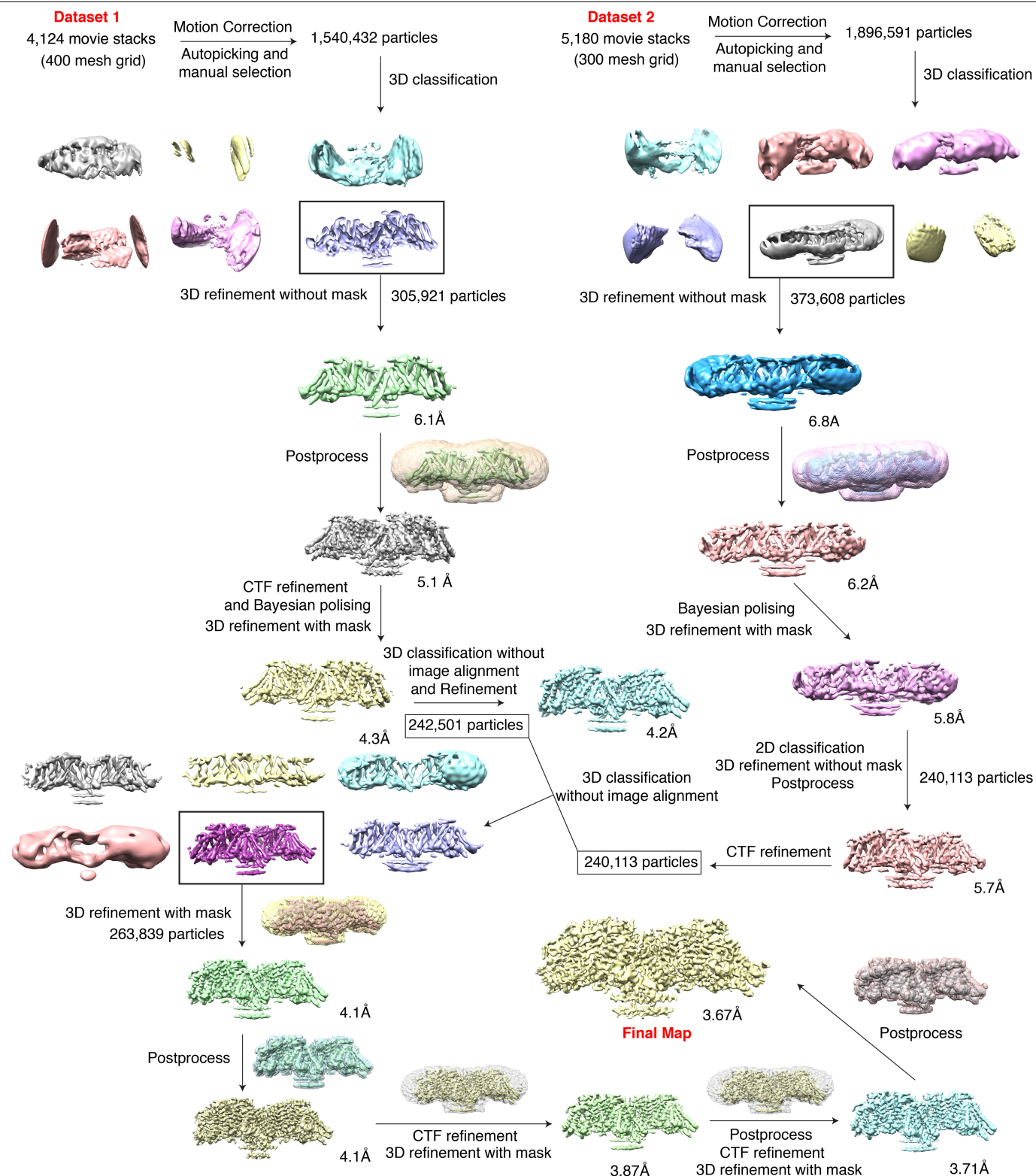
is shown on the right, and ACAT1 is indicated with an arrow. The experiment was reproduced three times on separate occasions with similar results.



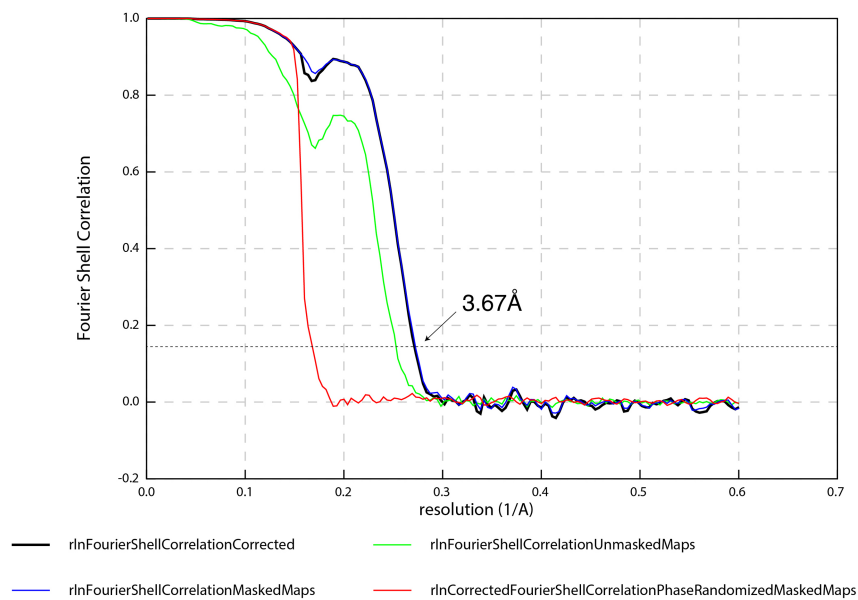
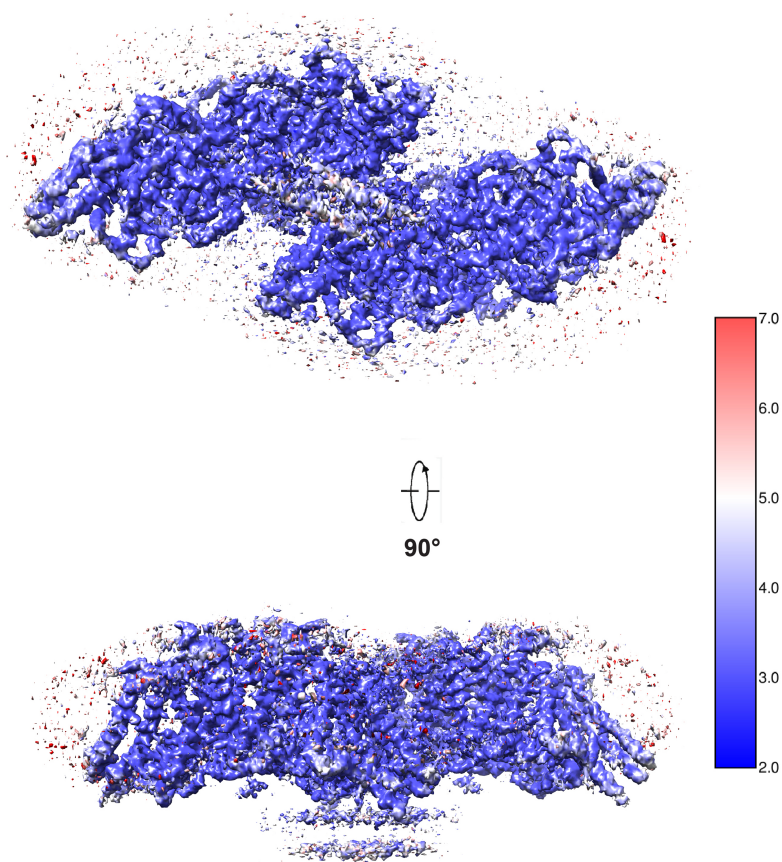
Extended Data Fig. 3 | Processing of data acquired from 200 kV cryo-TEM.

a, Representative gel-filtration chromatogram (Superose 6 Increase 10/30 column) of cross-linked ACAT1. The SDS-PAGE gel (with molecular weight markers) is shown on the right, and ACAT1 is indicated with an arrow. The

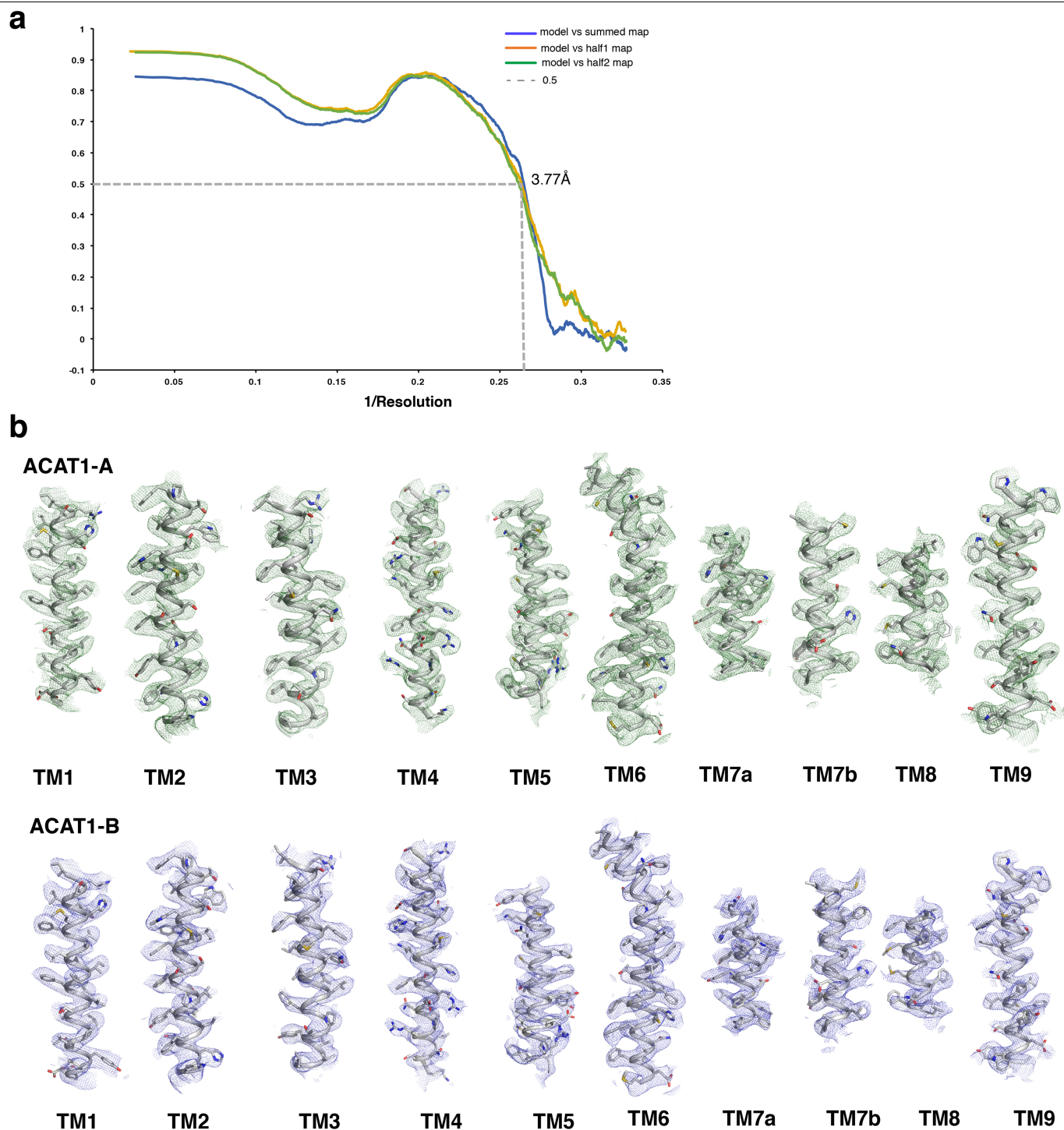
experiment was reproduced twice on separate occasions with similar results. **b**, The data processing workflow in RELION-3. The 3D classes and refinement results from the cryo-EM data are shown.



Extended Data Fig. 4 | Processing of data acquired from 300 kV cryo-TEM. The data processing workflow in RELION-3. The cryo-EM 3D classes as well as the masks used for the refinement are shown. After RELION-3 refinement, the final cryo-EM map was sharpened using post-process with a B-factor value of -150 \AA^2 .

a**b**

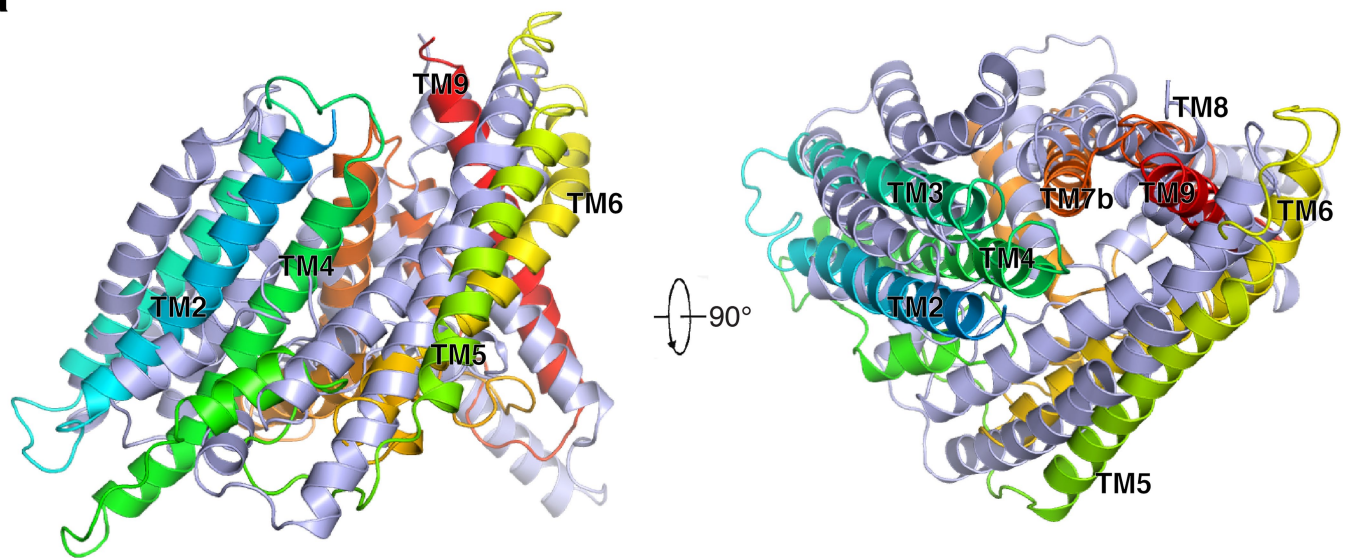
Extended Data Fig. 5 | FSC curve and estimation of the local resolution. a, FSC curve as a function of resolution using RELION-3 output. **b,** Density maps of the structure of ACAT1, coloured by local resolution estimation using ResMap.



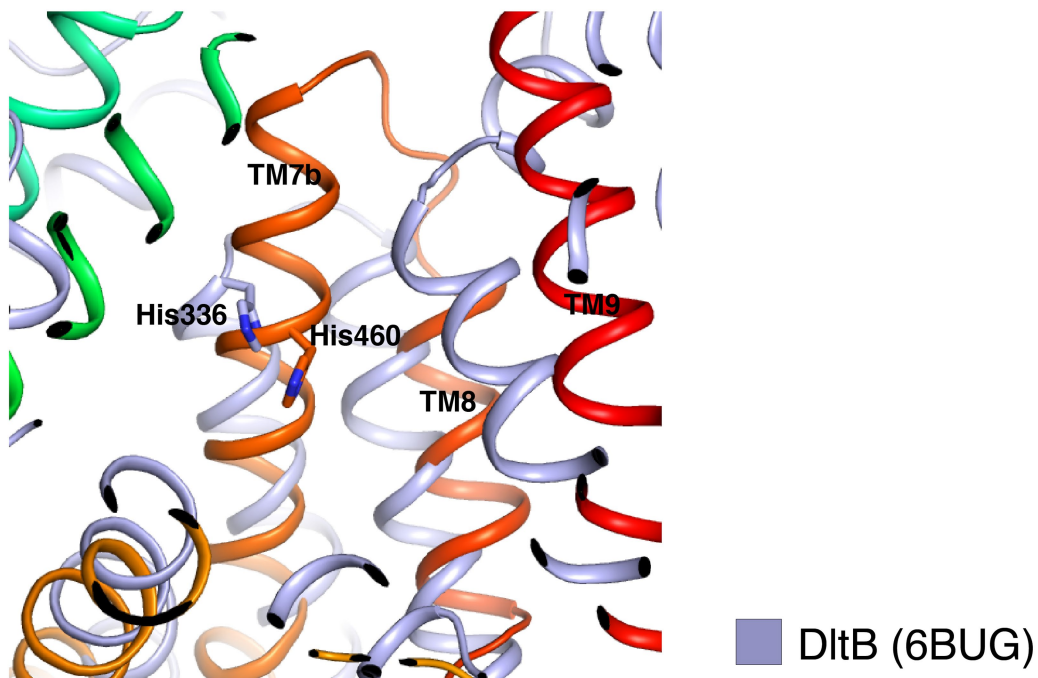
Extended Data Fig. 6 | Cryo-EM map of the structural elements of ACAT1.
a, The FSC curves calculated between the refined structure model and the half map used for refinement (orange), the other half map (green) and the full map

(blue). **b**, The major helices of ACAT1-A and ACAT1-B. The density map and model of the complex are shown as mesh and cartoons, respectively. Cryo-EM maps are contoured at the 5σ level.

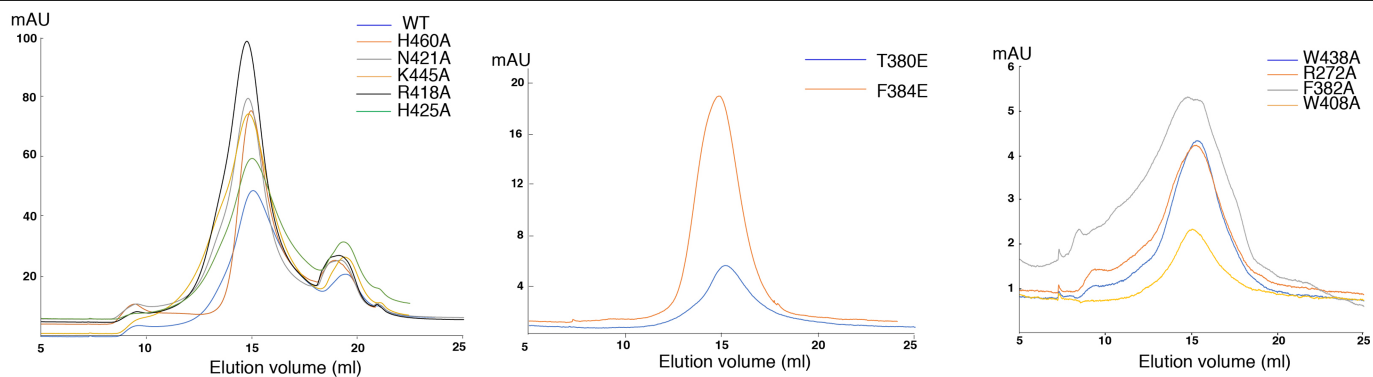
a



b

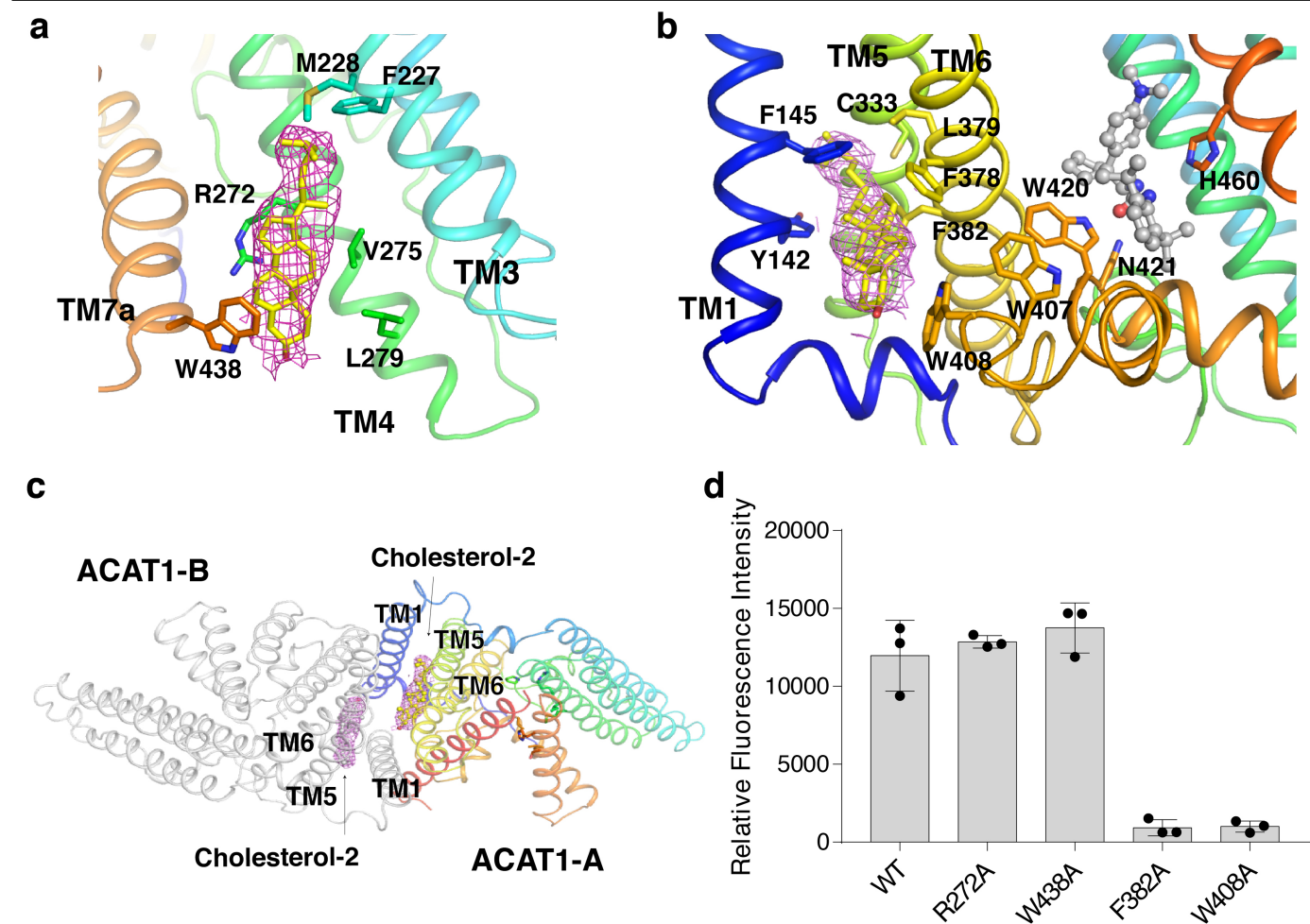


Extended Data Fig. 7 | Structural comparison of ACAT1 and DltB. **a**, Structural comparison from parallel to the membrane view (left) and from the lumen view (right). **b**, Structural comparison of the catalytic sites of ACAT1 and DltB. The catalytic histidine residues are shown as sticks.



Extended Data Fig. 8 | Purification of human ACAT1 variants for activity assays. Representative gel-filtration chromatograms (Superose 6 Increase 10/30 column) of ACAT1 variants in buffer containing 20 mM HEPES, pH 7.5,

150 mM NaCl and 1% CHAPS. The experiment was reproduced twice on separate occasions with similar results.



Extended Data Fig. 9 | The putative allosteric sites of ACAT1. a, b, Putative allosteric sites 1 (a) and 2 (b). Cholesterol (yellow) may form hydrophobic contacts that result in the stabilization of TM4 and TM6, respectively. **c,** Putative allosteric site 2 in the ACAT1 dimer. The cryo-EM maps of cholesterol

are contoured at the 5σ level. **d,** Functional validation of the putative allosteric sites. Data are mean \pm s.d. ($n=3$ technically independent experiments). Each experiment was reproduced at least twice on separate occasions with similar results.

Extended Data Table 1 | Cryo-EM data collection, refinement and validation statistics

	ACAT1-nevanimibe (EMDB-21390) (PDB 6VUM)
Data collection and processing	
Magnification	60024
Voltage (kV)	300
Electron exposure (e ⁻ /Å ²)	60
Defocus range (μm)	-1.0~-2.0
Pixel size (Å)	0.833
Symmetry imposed	C2
Initial particle images (no.)	3,437,023
Final particle images (no.)	263,839
Map resolution (Å)	3.67
FSC threshold	0.143
Map resolution range (Å)	3.0~5.0
Refinement	
Initial model used (PDB code)	<i>ab initio</i>
Model resolution (Å)	3.77
FSC threshold	0.5
Model resolution range (Å)	n/a
Map sharpening <i>B</i> factor (Å ²)	-150
Model composition	
Non-hydrogen atoms	14,388
Protein residues	1,636
Ligands	22
<i>B</i> factors (Å ²)	
Protein	115.83
Ligand	74.13
R.m.s. deviations	
Bond lengths (Å)	0.008
Bond angles (°)	1.191
Validation	
MolProbity score	1.80
Clashscore	4.15
Poor rotamers (%)	1.12
Ramachandran plot	
Favored (%)	89.31
Allowed (%)	10.69
Disallowed (%)	0

Reporting Summary

Nature Research wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Research policies, see [Authors & Referees](#) and the [Editorial Policy Checklist](#).

Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

n/a Confirmed

- ☒ ☐ The exact sample size (n) for each experimental group/condition, given as a discrete number and unit of measurement
- ☒ ☐ A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
- ☒ ☐ The statistical test(s) used AND whether they are one- or two-sided
Only common tests should be described solely by name; describe more complex techniques in the Methods section.
- ☒ ☐ A description of all covariates tested
- ☒ ☐ A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
- ☒ ☐ A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)
- ☒ ☐ For null hypothesis testing, the test statistic (e.g. F , t , r) with confidence intervals, effect sizes, degrees of freedom and P value noted
Give P values as exact values whenever suitable.
- ☒ ☐ For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
- ☒ ☐ For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
- ☒ ☐ Estimates of effect sizes (e.g. Cohen's d , Pearson's r), indicating how they were calculated

Our web collection on [statistics for biologists](#) contains articles on many of the points above.

Software and code

Policy information about [availability of computer code](#)

Data collection

SerialEM 3.7.10

Data analysis

CTFFIND4 4.1.8, Relion 2.1, BFactor 1.04, Refmac 5.8.0091, COOT 0.8.8, Phenix 1.13, PyMOL 1.8.6.0, MolProbity 4.3, Excel 16.13, Prism 7.00, Chimera 1.12, MotionCor2 1.2.1

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors/reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Research [guidelines for submitting code & software](#) for further information.

Data

Policy information about [availability of data](#)

All manuscripts must include a [data availability statement](#). This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A list of figures that have associated raw data
- A description of any restrictions on data availability

The 3D cryo-EM density map has been deposited in the Electron Microscopy Data Bank under the accession numbers EMD-21390. Atomic coordinate for the atomic model has been deposited in the Protein Data Bank under the accession numbers 6VUM. All other data is available from the corresponding authors upon reasonable request.

Field-specific reporting

Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

☒ Life sciences ☐ Behavioural & social sciences ☐ Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see [nature.com/documents/nr-reporting-summary-flat.pdf](https://www.nature.com/documents/nr-reporting-summary-flat.pdf)

Life sciences study design

All studies must disclose on these points even when the disclosure is negative.

Sample size	Sample size is not applicable for this study. Three independent experiments (n=3) were carried out for all the functional assays.
Data exclusions	No data was excluded.
Replication	Each experiment was reproduced at least two times on separate occasions. Experimental findings were reliably reproduced.
Randomization	Randomization was not necessary as the independent variables to be tested were sufficient for the functional interpretations within this study. i.e. WT vs mutant vs control conditions or dose response determination.
Blinding	Blinding is not necessary or valid for the purposes of structural determination. For functional analysis, blinding was not necessary due to the quantitative nature of the experiment. All experimental data acquired is included in our statistical analysis.

Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

Materials & experimental systems

n/a	Involved in the study
<input checked="" type="checkbox"/>	<input type="checkbox"/> Antibodies
<input type="checkbox"/>	<input checked="" type="checkbox"/> Eukaryotic cell lines
<input checked="" type="checkbox"/>	<input type="checkbox"/> Palaeontology
<input checked="" type="checkbox"/>	<input type="checkbox"/> Animals and other organisms
<input checked="" type="checkbox"/>	<input type="checkbox"/> Human research participants
<input checked="" type="checkbox"/>	<input type="checkbox"/> Clinical data

Methods

n/a	Involved in the study
<input checked="" type="checkbox"/>	<input type="checkbox"/> ChIP-seq
<input checked="" type="checkbox"/>	<input type="checkbox"/> Flow cytometry
<input checked="" type="checkbox"/>	<input type="checkbox"/> MRI-based neuroimaging

Eukaryotic cell lines

Policy information about [cell lines](#)

Cell line source(s)	The protein were expressed from HEK-293S GnTI- (ATCC).
Authentication	No further authentication was performed for commercially available cell lines.
Mycoplasma contamination	periodically test negative
Commonly misidentified lines (See ICLAC register)	None of the cell lines used is listed in the database of commonly misidentified cell lines maintained by ICLAC.

Publisher Correction: Enhanced ferroelectricity in ultrathin films grown directly on silicon

<https://doi.org/10.1038/s41586-020-2297-6>

Correction to: *Nature* <https://doi.org/10.1038/s41586-020-2208-x>

Published online 22 April 2020



Check for updates

Suraj S. Cheema, Daewoong Kwon, Nirmaan Shanker, Roberto dos Reis, Shang-Lin Hsu, Jun Xiao, Haigang Zhang, Ryan Wagner, Adhiraj Datar, Margaret R. McCarter, Claudy R. Serrao, Ajay K. Yadav, Golnaz Karbasian, Cheng-Hsiang Hsu, Ava J. Tan, Li-Chen Wang, Vishal Thakare, Xiang Zhang, Apurva Mehta, Evguenia Karapetrova, Rajesh V Chopdekar, Padraic Shafer, Elke Arenholz, Chenming Hu, Roger Proksch, Ramamoorthy Ramesh, Jim Ciston & Sayeef Salahuddin

In the first sentence of this Article, “Ultrathin ferroelectric materials could potentially enable low-power **perovskite ferroelectric tetragonality** logic and nonvolatile memories”, the words “perovskite ferroelectric tetragonality” should not have appeared. This error has been corrected online.

Publisher Correction: Intensive farming drives long-term shifts in avian community composition

<https://doi.org/10.1038/s41586-020-2310-0>

Correction to: *Nature* <https://doi.org/10.1038/s41586-020-2090-6>

Published online 18 March 2020



Check for updates

J. Nicholas Hendershot, Jeffrey R. Smith, Christopher B. Anderson, Andrew D. Letten, Luke O. Frishkoff, Jim R. Zook, Tadashi Fukami & Gretchen C. Daily

In the 'Data availability' section of this Article, owing to an error during the production process, the DOI was incorrectly shown as '<https://doi.org/10.6084/m9.figshare.6025748>' instead of '<https://doi.org/10.6084/m9.figshare.11366201>'. The original Article has been corrected online.

Editorial Expression of Concern: Exploring the quantum speed limit with computer games

<https://doi.org/10.1038/s41586-020-2274-0>

Addendum to: *Nature* <https://doi.org/10.1038/nature17620>

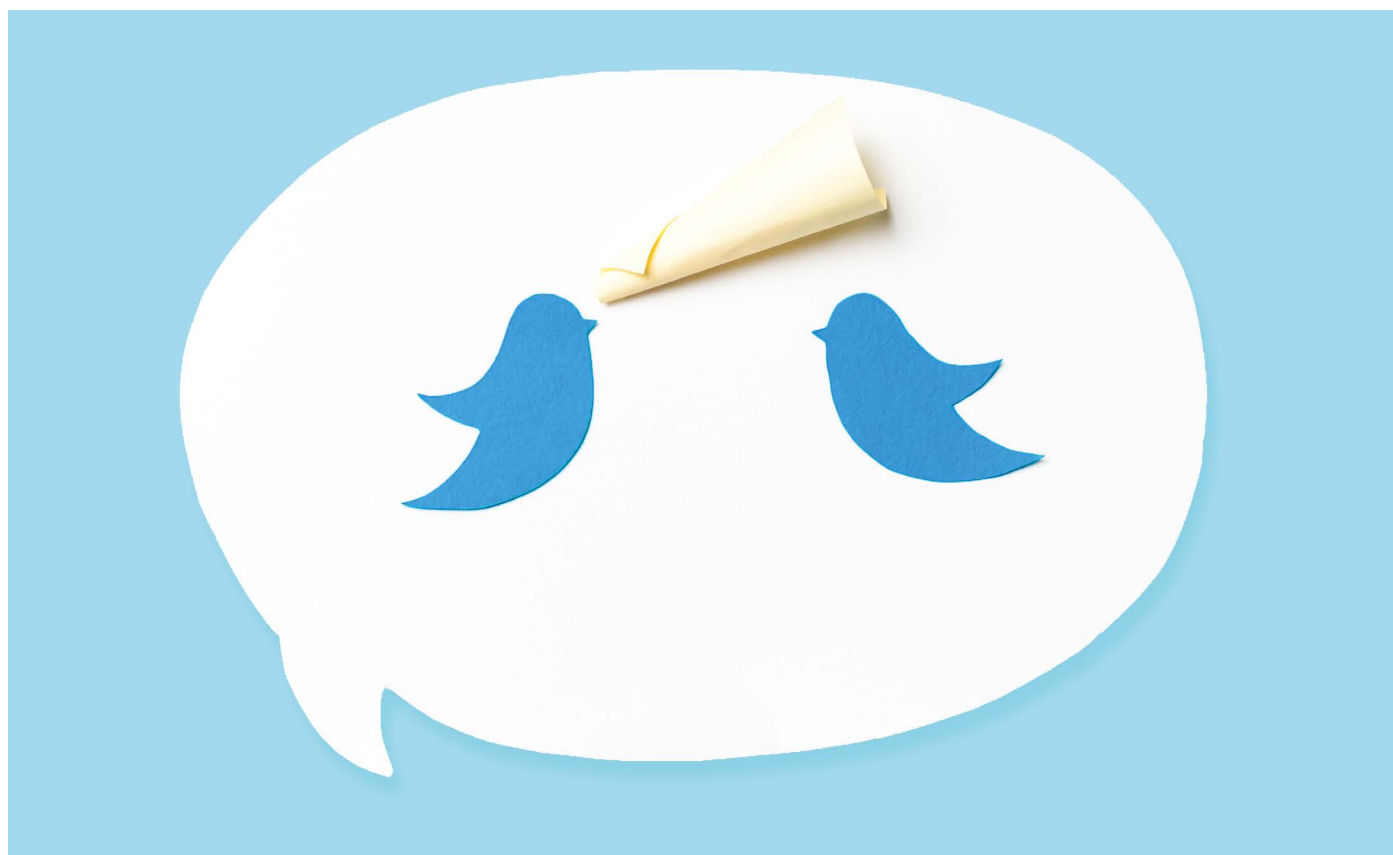
Published online 13 April 2016



Check for updates

Jens Jakob W. H. Sørensen, Mads Kock Pedersen, Michael Munch, Pinja Haikka, Jesper Halkjær Jensen, Tilo Planke, Morten Ginnerup Andreasen, Mirosław Gajdacz, Klaus Mølmer, Andreas Lieberoth & Jacob F. Sherson

The authors have alerted the editors of *Nature* to an error in the code underlying the work in this Letter, and have informed us that this error will have an impact on the conclusions that can reliably be drawn. *Nature* is working with the authors to resolve the matter, but in the meantime, readers are cautioned against using results from this Letter.



ADAPTED FROM GETTY

FIGHT CORONAVIRUS MISINFORMATION

Use social media to spread good pandemic science.

By Samantha Yammine

Unsurprisingly, millions of people are talking about the coronavirus on social media. According to the analytics platform Sprinklr, there were more than 19 million mentions of coronavirus across social media on 11 March (the day the World Health Organization called the outbreak a pandemic), and a report by Twitter in early April said that COVID-19-related tweets were being shared every 45 milliseconds.

As the pandemic continues to evolve and we look towards long-term management strategies, we must continue these conversations – and make sure that science is a part of them. But, as science-communication experts have been saying, simply spewing scientific facts from a soapbox isn't enough: research shows

that it's more important to start a dialogue.

When used strategically, social media can make it easier to have such conversations at scale. I am a scientist who shares online updates about COVID-19, and my coronavirus social-media content has been viewed millions of times. I feel that I'm doing more than adding

“Acknowledging that it isn't easy to be cooped up at home is not only honest, but relatable. And relatability is a key component of trust.”

sound to all the noise because my engagement rates have been as high as 24% on Instagram, more than 10 times the industry standard, and 49 times the standard on my COVID-19 tweets.

And it's particularly important to share accessible science through social media because trolls and conspiracy theorists are spreading seeds of doubt and misinformation that can have dangerous consequences.

Good science communication involves storytelling, avoiding jargon and making science accessible. Here are tips on sharing information at scale on social media.

Amplify first. Not everyone has the time or skillset to create new material for sharing on social media, but amplifying the messages of

Had a good,
bad or
ugly career
experience?
Tell the
world.

Work with one of our
editors to develop your
article. Email us at
naturecareerseditor@
nature.com

nature
careers

A80550

Work / Careers

others is a helpful way to contribute. Your likes, shares and retweets are a form of social currency: instead of angrily sharing only things you disagree with, use your currency to boost credible work to help good content to go viral.

If you do want to make content, aim to fill a gap: address a common point of confusion no one has tackled yet, or find a unique way to share knowledge (for example, through art, dance, rap or pop-culture references).

Avoid 'hot takes'. The scientific process is rooted in a culture of debate and collaborative criticism – we're all used to attacking things. Before firing off a juicy tweet, think deeply about whether it'll cause more confusion than clarity if seen by people who aren't your colleagues. If a minor, debatable detail wouldn't change public-health guidelines, perhaps it's an argument best saved for another time.

Create content tailored to your target audience. Make your content as data driven as your research is. Perhaps you want to reach young people who are disregarding social distancing. Your next step would be to see what platform that demographic uses, and how they communicate there. For example, 65% of Instagram users globally are under the age of 34; 72% of teens who are online are using Instagram; and 41% of people who use the video-sharing platform TikTok are between 16 and 24 years old.

But simply a clip re-shared from the evening news isn't likely to spread far on TikTok. At any given time, there will be trending audio clips, dances, challenges and memes, and new videos shared with these trends often enjoy a lot of traffic. Repackaging scientific information in trending formats can make it more visible.

Have a hook, but avoid clickbait. A standard scientific paper builds information through a few paragraphs in the introduction, with the 'hook' or main findings in the final sentence of that section. To reach a broader audience, reverse that process: start with that hook to entice readers to delve into the details.

Hooks framed as questions often do well, but readers can be disappointed if the answer is a simple yes or no. And leading questions that imply doubt can spread misinformation if people don't read on. So instead of 'Does hand sanitizer work only on bacteria?', try 'How does hand sanitizer kill viruses and bacteria?'

A picture or graphic can serve as an exciting hook to make someone pause while scrolling through their newsfeed. Use design tools such as BioRender, Canva and VSCO to create and edit photos and graphics.

On Instagram and TikTok, use hashtags to reach fresh audiences. Users can discover content through search functions on social-media platforms, so you can drive traffic to your post without having many followers by

using hashtags related to COVID-19. Twitter indexes content in search results by any word in a tweet, but users can search content on Instagram and TikTok only using hashtags.

On Instagram, you can use up to 30 hashtags per post: use them all! Include a mix of broad, high-traffic hashtags such as #science, and niche hashtags such as #microbiology, #educational or #ScienceTeacher.

On TikTok, less is more: check out the latest hashtag trend and see if you can find a way to fit the information you want to share into that video format. Although you might not be able to pack much information into the one minute allowed per video, you can get new conversations started in the comments section.

Your hashtags should always be relevant to the content you're sharing, but you can broaden your reach beyond science 'echo chambers' (content that preaches to the converted, reinforcing existing beliefs) by incorporating other trends. Creative visuals that showcase science through a popular aesthetic – for example, by sharing your information alongside #calligraphy, #fashion or even #pastels – can put your content in front of high-traffic target communities.

And using more colloquial names for scientific terms, such as #covid or #corona, can help to broaden your reach beyond scientists.

If you're going to bust myths, do it compassionately. People are more likely to listen to someone who listens to them. To avoid being dismissive, I usually say, "I understand why you're nervous about ____, this is a really scary time. But ____." UK science-funder Wellcome's latest *Global Monitor*, published in June 2019, found that 18% of people have a high level of trust in scientists, and 54% have a medium level. We will get a lot further by fostering trusting relationships.

Be your authentic self. Scientists are people too. We are more than our graphs: epidemic curves showing the impact of physical distancing are important, but acknowledging that it isn't easy to be cooped up at home is not only honest, but relatable. And relatability is a key component of trust.

Including yourself in your story can help to convey warmth and trust, and to get people to listen and take your message to heart.

No public-health research is complete until the key findings are effectively communicated and, ideally, implemented. Although the scale of online platforms adds challenges to this task, it can be leveraged to share conversations about the life-saving science we need most.

Samantha Yammine received her PhD from the University of Toronto in Canada studying neural stem cells, and is an independent science communicator known online as Science Sam online.



Where I work Anne-Marie Coriat

As head of Wellcome's UK and Europe Research Landscape, I look at how the global research system comes together in the United Kingdom and internationally. My job focuses on how well everything joins up.

I get a lot of energy from discussing issues with a variety of people. We recently completed a three-year project looking at how we fund PhD training, and learnt that scientists are concerned by shortcomings in research culture. We're now promoting programmes that demonstrate both scientific excellence and support for student development.

The Wellcome building is delightful. Its ground floor, known as 'the Street', is devoted to cafe-style tables, and the buzz can be very helpful when I am getting my thoughts together.

But when I need uninterrupted time to think, I head next door to the Reading Room, a library and gallery space in the Wellcome Collection, our free museum and library. The objects and books on display mix art, science and history in a calm environment

that invites you to think, to challenge your own assumptions. It is a welcoming space, a great combination of natural light and soft tones. When I've been there, it feels like I've had a break even though I'm still working.

I need a combination of quietness and noise, and during this three-year project on PhD training, I've had to seek out these quiet spaces to work, and then get out and test the ideas on people.

Since the UK lockdown in response to the new coronavirus, our Wellcome buildings have been closed. But in many ways, we are busier than ever, tackling the pandemic through coalitions such as COVID-Zero; this calls on businesses to help raise US\$1 billion for pandemic research and development, testing, treatment and vaccines. When the time is right, I will enjoy my next visit to the Wellcome library – but, for now, I am lucky to be able to work remotely.

Anne-Marie Coriat is Wellcome's head of UK and Europe Research Landscape in London. **Interview by James Mitchell Crow.**

Photographed for *Nature* by
Leonora Saunders.